NAVAL POSTGRADUATE SCHOOL
Monterey, California

THESIS

SPECTRUM ESTIMATION USING
EXTRAPOLATED TIME SERIES

by

Robert Timothy Thornlow

December 1990

Thesis Advisor: Ralph Hippenstiel

UNCLASSIFIED
SECURITY CLASSIFICATION OF THIS PAGE

REPORT DOCUMENTATION PAGE

Form Approved OMB No. 0704-0188

1a REPORT SECURITY CLASSIFICATION UNCLASSIFIED	1b RESTRICTIVE MARKINGS
2a SECURITY CLASSIFICATION AUTHORITY	3 DISTRIBUTION / AVAILABILITY OF REPORT Approved for public release; distribution is unlimited
2b DECLASSIFICATION / DOWNGRADING SCHEDULE	
4 PERFORMING ORGANIZATION REPORT NUMBER(S)	5 MONITORING ORGANIZATION REPORT NUMBER(S)
6a NAME OF PERFORMING ORGANIZATION Naval Postgraduate School 6b OFFICE SYMBOL (If applicable) EC	7a NAME OF MONITORING ORGANIZATION Naval Postgraduate School
6c ADDRESS (City, State, and ZIP Code) Monterey, CA 93943-5000	7b ADDRESS (City, State, and ZIP Code) Monterey, CA 93943-5000
8a NAME OF FUNDING / SPONSORING ORGANIZATION 8b OFFICE SYMBOL (If applicable)	9 PROCUREMENT INSTRUMENT IDENTIFICATION NUMBER
8c ADDRESS (City, State, and ZIP Code)	10 SOURCE OF FUNDING NUMBERS PROGRAM ELEMENT NO / PROJECT NO / TASK NO / WORK UNIT ACCESSION NO

11 TITLE (Include Security Classification)
SPECTRUM ESTIMATION USING EXTRAPOLATED TIME SERIES

12 PERSONAL AUTHOR(S)
THORNLOW, Robert Timothy

13a TYPE OF REPORT	13b TIME COVERED	14 DATE OF REPORT (Year, Month, Day)	15 PAGE COUNT
Master's Thesis	FROM _____ TO _____	1990 December	106

16 SUPPLEMENTARY NOTATION The views expressed in this thesis are those of the author and do not reflect the official policy or position of the Department of Defense or the US Government.

17 COSATI CODES			18 SUBJECT TERMS (Continue on reverse if necessary and identify by block number)
FIELD	GROUP	SUB-GROUP	data extrapolation; periodogram; AR spectral estimates

19 ABSTRACT (Continue on reverse if necessary and identify by block number)

Several techniques for estimation of the power spectral density (PSD), or simply the spectrum, of discretely sampled short data sequences are investigated in this thesis. These techniques are directly based on the fast Fourier transform (FFT) of the data sequence. However, an extrapolated version of the original data sequence will be used instead of the original. The aim is to enhance the resolution of the true spectral components of the signal without adding any false frequency components. Several data extrapolation models are analyzed and their performance at different signal to noise ratios (SNR's) and model orders are examined.

20 DISTRIBUTION / AVAILABILITY OF ABSTRACT	21 ABSTRACT SECURITY CLASSIFICATION	
☒ UNCLASSIFIED/UNLIMITED ☐ SAME AS RPT ☐ DTIC USERS	UNCLASSIFIED	
22a NAME OF RESPONSIBLE INDIVIDUAL HIPPENSTIEL, Ralph	22b TELEPHONE (Include Area Code) 408-646-2768	22c OFFICE SYMBOL EC/Hi

DD Form 1473, JUN 86 *Previous editions are obsolete.* SECURITY CLASSIFICATION OF THIS PAGE

S/N 0102-LF-014-6603 UNCLASSIFIED

Approved for public release; distribution is unlimited

Spectrum Estimation Using Extrapolated Time Series

by

Robert Timothy Thornlow
Lieutenant, United States Navy
B.S.E.E., North Carolina State University, 1984

Submitted in partial fulfillment of the requirements for the degree of

MASTER OF SCIENCE IN ELECTRICAL ENGINEERING

From the

NAVAL POSTGRADUATE SCHOOL
December 1990

Authors: Robert Timothy Thornlow
Robert Timothy Thornlow

Approved by: Ralph Hippenstiel
Ralph Hippenstiel, Thesis Advisor

Murali Tummala
Murali Tummala, Second Reader

Michael A. Morgan
Michael A. Morgan, Chairman
Department of Electrical and Computer Engineering

ABSTRACT

Several techniques for estimation of the power spectral density (PSD), or simply the spectrum, of discretely sampled short data sequences are investigated in this thesis. These techniques are directly based on the fast Fourier transform (FFT) of the data sequence. However, an extrapolated version of the original data sequence is used instead of the original. The aim is to enhance the resolution of the true spectral components of the signal without adding any false frequency components. Several data extrapolation models are analyzed and their performance at different signal to noise ratios (SNRs) and model orders are examined.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

I. INTRODUCTION

This thesis investigates several techniques for enhancing the spectrum estimation of short data sequences. A short data sequence is defined as one in which the spectral resolution required is of the same order, or better, as the reciprocal sequence length. The techniques to be investigated will involve extrapolating the original data sequence and then performing a basic periodogram on this new data sequence. By extending the original data sequence we hope to enhance the true spectral components and the spectral resolution of the signal without adding any false spectral components.

The method of spectrum estimation to be used on all extended data sequences is the basic periodogram. The periodogram is defined as

$$\hat{P}_x\left(\varepsilon^{j\omega}\right) = \frac{1}{M}\left|\sum_{n=0}^{M-1} x[n]\varepsilon^{-j\omega n}\right|^2 \tag{1.1}$$

where x[n] is the extrapolated data sequence and M is the number of points of the extended data sequence.

This thesis is arranged into 5 chapters and 4 appendices. Chapter II is devoted to data extrapolation via linear predictive filtering. In Chapter III the data is modeled as the impulse response of an infinite impulse response (IIR) filter. Data extrapolation by dominant eigenvectors is discussed in Chapter IV. In the last chapter some general conclusions of the work carried out in this thesis are presented, and suggestions for future investigations are given. Appendix A covers several other methods of AR parameter estimation. Appendix B gives some results from a causal Prony's method of

extrapolation. Appendix C contains some results from some miscellaneous test sequences while computer simulation programs are included in Appendix D.

II. LINEAR PREDICTIVE FILTERING

A. DEFINITION

In this chapter, all original test data sequences will be extrapolated by linear prediction. In linear prediction, a given set of sample values of a random sequence x is used to predict the values of the sequence some time into the future. We assume that only P previous values of the sequence are used in the prediction. Now the problem is to find the prediction coefficients $-a_1^*, -a_2^*, \ldots, -a_P^*$ to produce an estimate of the current value of the random process x[n] from P previous values. The estimate is written as

$$\hat{x}[n] = -a_1^* x[n-1] - a_2^* x[n-2] - \ldots - a_p^* x[n-P] \tag{2.1}$$

where * represents the complex conjugation.

The error in the estimate is given by the difference

$$e[n] = x[n] - \hat{x}[n] = \sum_{k=0}^{P} a_k^* x[n-k] \tag{2.2}$$

where a_0 is defined as 1.

The approach is to choose the prediction filter coefficients to minimize the mean-square error

$$\sigma_e^2 = E\left[|e[n]|^2\right] = E\left[|x[n] - \hat{x}[n]|^2\right]. \tag{2.3}$$

To find the optimal prediction filter coefficients we apply the Orthogonality Theorem [Ref. 1:p. 308] which states that

$$E[x[n-i]e^*[n]] = 0 \quad \text{for i} = 1,2,\ldots P \tag{2.4}$$

or

$$r_{xx}[i] = -\sum_{\ell=1}^{P} a_\ell r_{xx}[i-\ell] \quad \text{for i} = 1,2,\ldots,P \tag{2.5}$$

where r_{xx} is the autocorrelation function of x.

The minimum prediction error is found to be

$$e_{\min} = r_{xx}[0] + \sum_{k=1}^{P} a_k r_{xx}[-k]. \tag{2.6}$$

These equations are identical to the Yule-Walker equations for autoregressive (AR) processes [Ref. 2:p. 157]. Therefore, the optimal linear prediction coefficients are just the AR parameters, when the order of the AR process is identical to the order of the linear predictor.

B. AR PARAMETER ESTIMATION

For good estimates of the AR parameters or prediction coefficients, a maximum likelihood estimator (MLE) is usually employed. Several methods were tried in this thesis including the autocorrelation method, the covariance method, the forward-backward method (modified covariance method), the modified forward-backward method, and the Burg method. The two methods which proved to be the most robust were the forward-backward method and the modified forward-backward method. Explanations and results from the other methods are given in Appendix A.

1. Forward-backward Method (Modified Covariance)

For an AR process the optimal forward predictor is given by Eq. (2.1). The corresponding optimal backward predictor is given by [Ref. 2:p. 226]

$$\hat{x}[n] = \sum_{k=1}^{P} -a_k x[n+k]. \tag{2.7}$$

The forward-backward method estimates the prediction coefficients by minimizing the average of the estimated forward and backward prediction error powers. The average of the estimated forward and backward prediction error powers is given by

$$\hat{\rho} = \frac{1}{2}\left(\hat{\rho}^f + \hat{\rho}^b\right) \tag{2.8}$$

where the forward and backward prediction error powers are respectively defined as

$$\hat{\rho}^f = \frac{1}{N-P}\sum_{n=P}^{N-1}\left|x[n] + \sum_{k=1}^{P} a[k]x[n-k]\right|^2$$

$$\hat{\rho}^b = \frac{1}{N-P}\sum_{n=0}^{N-1-P}\left|x[n] + \sum_{k=1}^{P} a^*[k]x[n+k]\right|^2, \tag{2.9}$$

and N is the number of data points in the random sequence x. To minimize Eq. (2.8) we differentiate $\hat{\rho}$ with respect to the real and imaginary parts of a[k] for k=1,2,...P and set the expression equal to zero. After simplification this becomes

$$\sum_{k=1}^{P} a[k]\left(\sum_{n=P}^{N-1} x[n-k]x^*[n-\ell] + \sum_{n=0}^{N-1-P} x^*[n+k]x[n+\ell]\right)$$
$$= -\left(\sum_{n=P}^{N-1} x[n]x^*[n-\ell] + \sum_{n=0}^{N-1-P} x^*[n]x[n+\ell]\right) \tag{2.10}$$

for ℓ=1,2,...,P.

Defining

$$C_{xx}[j,k] = \frac{1}{2(N-P)}\left(\sum_{n=P}^{N-1} x^*[n-j]x[n-k] + \sum_{n=0}^{N-1-P} x[n+j]x^*[n+k]\right), \tag{2.11}$$

Eq. (2.10) can be written in matrix form as

$$\begin{bmatrix} C_{xx}[1,1] & C_{xx}[1,2] & \cdots & C_{xx}[1,P] \\ C_{xx}[2,1] & C_{xx}[2,2] & \cdots & C_{xx}[2,P] \\ \vdots & \vdots & \ddots & \vdots \\ C_{xx}[P,1] & C_{xx}[P,2] & \cdots & C_{xx}[P,P] \end{bmatrix} \begin{bmatrix} a[1] \\ a[2] \\ \vdots \\ a[P] \end{bmatrix} = -\begin{bmatrix} C_{xx}[1,0] \\ C_{xx}[2,0] \\ \vdots \\ C_{xx}[P,0] \end{bmatrix}. \tag{2.12}$$

The first matrix is a P-by-P autocorrelation matrix while the last matrix is called the cross correlation vector. The prediction coefficients are easily solved for from Eq. (2.12).

Intuitively, the order P should be as large as possible in order to have a large data base for the predictor. The use of too large a value of P can give rise to spurious spectral peaks in the AR spectral estimation using the modified covariance method [Ref. 3:p. 347]. For the best performance of the forward-backward linear predictor (FBLP) method, Lang [Ref. 4:p. 720] suggests the value

$$P \cong \frac{N}{3} \tag{2.13}$$

where once again N is the original data length.

Akaike [Ref. 5:pp. 203-217] proposes two other methods of model order selection based on the estimated prediction error power.

a. Test Data

Two types of test data were used for computer simulation of all methods. The first set was obtained from a paper by S. M. Kay and S. L. Marple [Ref. 6:p. 1380-1414]. The data is made up of three sinusoids and a colored noise process which is obtained by filtering a white Gaussian process. The three sinusoids are at fractional frequencies of the sampling frequency of 0.10, 0.20, and 0.21 and have SNR's of +10, +30, and +30 dB respectively. SNR is defined as the ratio of the sinusoid power to the total power in the passband noise process. The noise process passband is centered at 0.35 and has a bandwidth of 0.30. Figure 1 is a plot of this 64 point test sequence and the true PSD. Figure 2 is the periodogram of the original, nonextrapolated data sequence using a Hamming data window. A Hamming window is used in all of the periodogram estimations in this thesis.

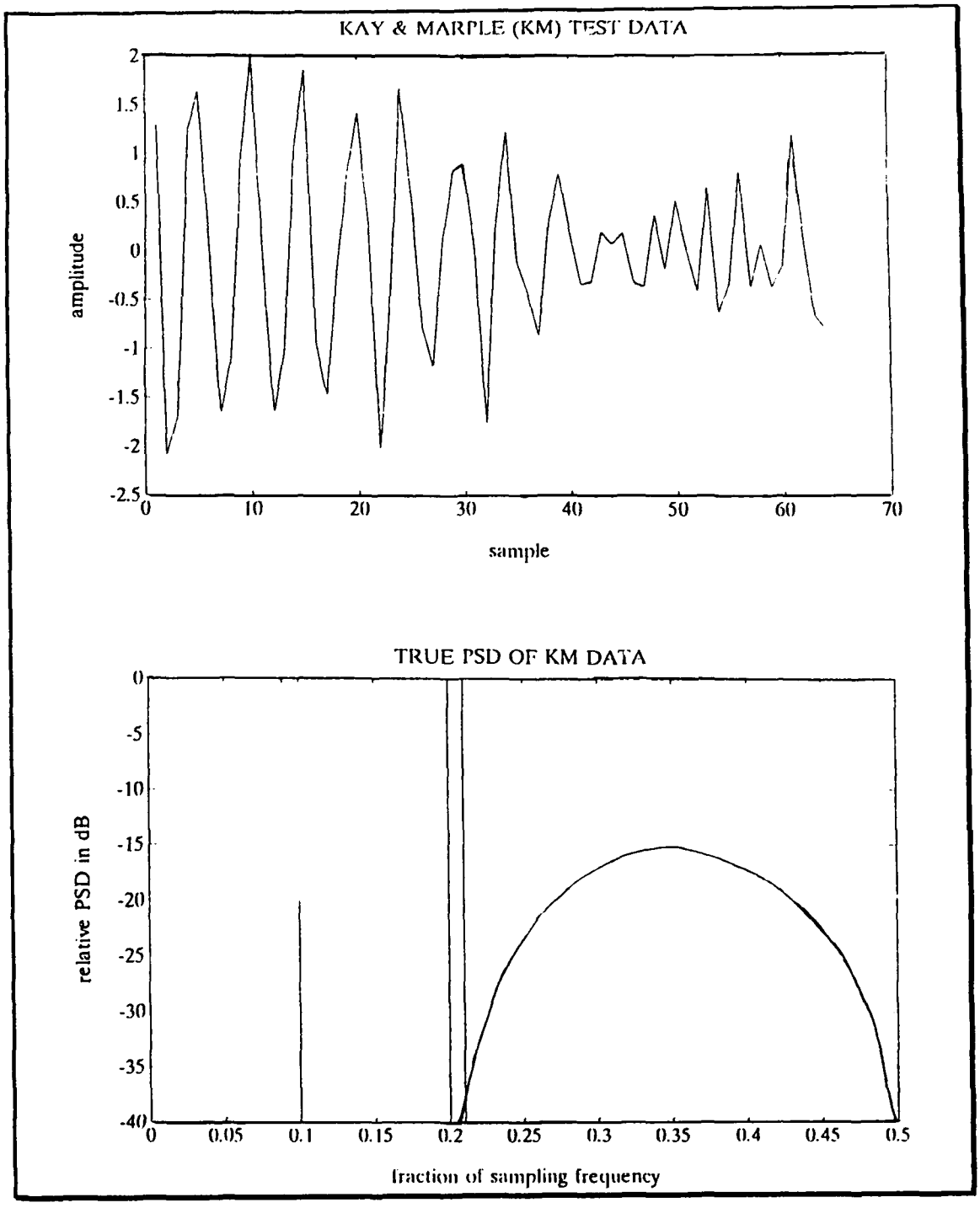

Figure 1. Kay and Marple Data Set with True Power Spectral Density

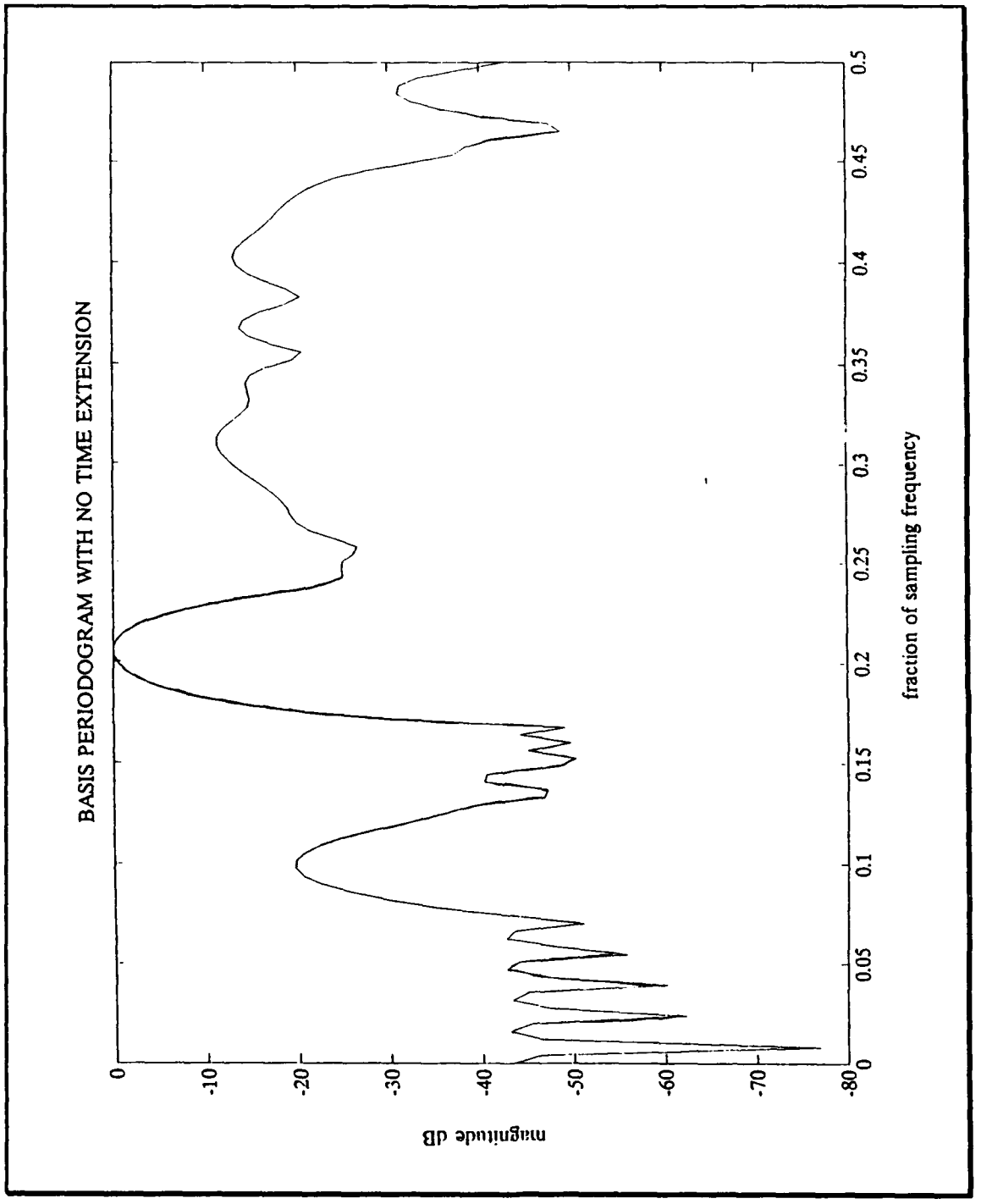

Figure 2. Periodogram of Kay and Marple Data (Hamming Window)

The second type of test data is made up of two sinusoids corrupted by white noise . This is also a 64 point test sequence. This type of test data will be composed of different frequencies, phases, and SNRs to see how each method of spectrum estimation resolves the sinusoidal frequencies. A sampling frequency of 1000 Hz is used to generate the signals, therefore the maximum frequency of any single sinusoid is 500 Hz. Using ordinary FFT routines on non-extended 64 point data sequences, the spectral resolution can be as good as 0.8(1/64msec)=12.5 Hz. Using extended data sequences, it will be shown that the resolution can be improved to 2.0 Hz for an appropriate SNR.

For comparison purposes, the periodograms of the extrapolated data sequences are plotted with the autoregressive (AR) spectrum estimations of the same data sequences. The AR spectrum is defined as

$$S_{AR}\left(\varepsilon^{j\omega}\right) = \frac{1}{\left|A\left(\varepsilon^{j\omega}\right)\right|^2} \tag{2.13}$$

where

$$A\left(\varepsilon^{j\omega}\right) = \sum_{n=0}^{P} a[n]\varepsilon^{-j\omega n} \tag{2.14}$$

and of course the AR coefficients are identical to the prediction filter coefficients.

The dashed lines on each plot represent the positions of the true frequencies of the test data.

b. FBLP Results

Figure 3 is the extrapolated version of the Kay and Marple (KM) data. The data was extrapolated by using linear prediction as in Eq. (2.1). The prediction coefficients were obtained from the FBLP method utilizing Eq. (2.12). The data length was extrapolated to 320 points using a model order of 16.

Care must be taken when extrapolating the data sequence. If the correlation matrix in the calculation of the prediction coefficients is not positive definite, the resulting poles in the filter of the AR model are not guaranteed to lie within the unit circle. This could cause the extrapolation to grow without bound and will have devastating effects on the spectrum estimation. The forward-backward method does not guarantee a stable all pole filter, although in practice the extrapolation is usually bounded. We should check to ensure all poles of the AR process lie within the unit circle before we extrapolate the data.

Figure 4 shows the periodogram of the extrapolated KM data and the respective AR spectral estimation. Both methods correctly calculate the three sinusoidal frequencies present. The colored noise process is also given a fair representaion. When the data is extrapolated to 576 points, an even better PSD estimate is obtained from the FBLP extrapolation. Figure 5 shows the results.

For the two sinusoids corrupted with white noise, good results are obtained. Figure 6 shows the spectrums of two sinusoids with a 10 Hz separation and a SNR of 12 dB. The original data sequence was extrapolated to 320 points. Both sine waves have a 0 degree initial phase angle. The frequencies are correctly calculated using the periodogram of the FBLP extrapolated sequence while the AR spectral estimation is unable to distinguish the two closely spaced sinusoids.

When the initial phase angle between the two sine waves is changed to 45 degrees, the periodogram of the FBLP extrapolated sequence, Figure 7, distinguishes the two frequencies, even though one of the estimated spectral components is slightly off. The AR spectrum estimation is unable to resolve them. Once again the data was extrapolated to 320 points.

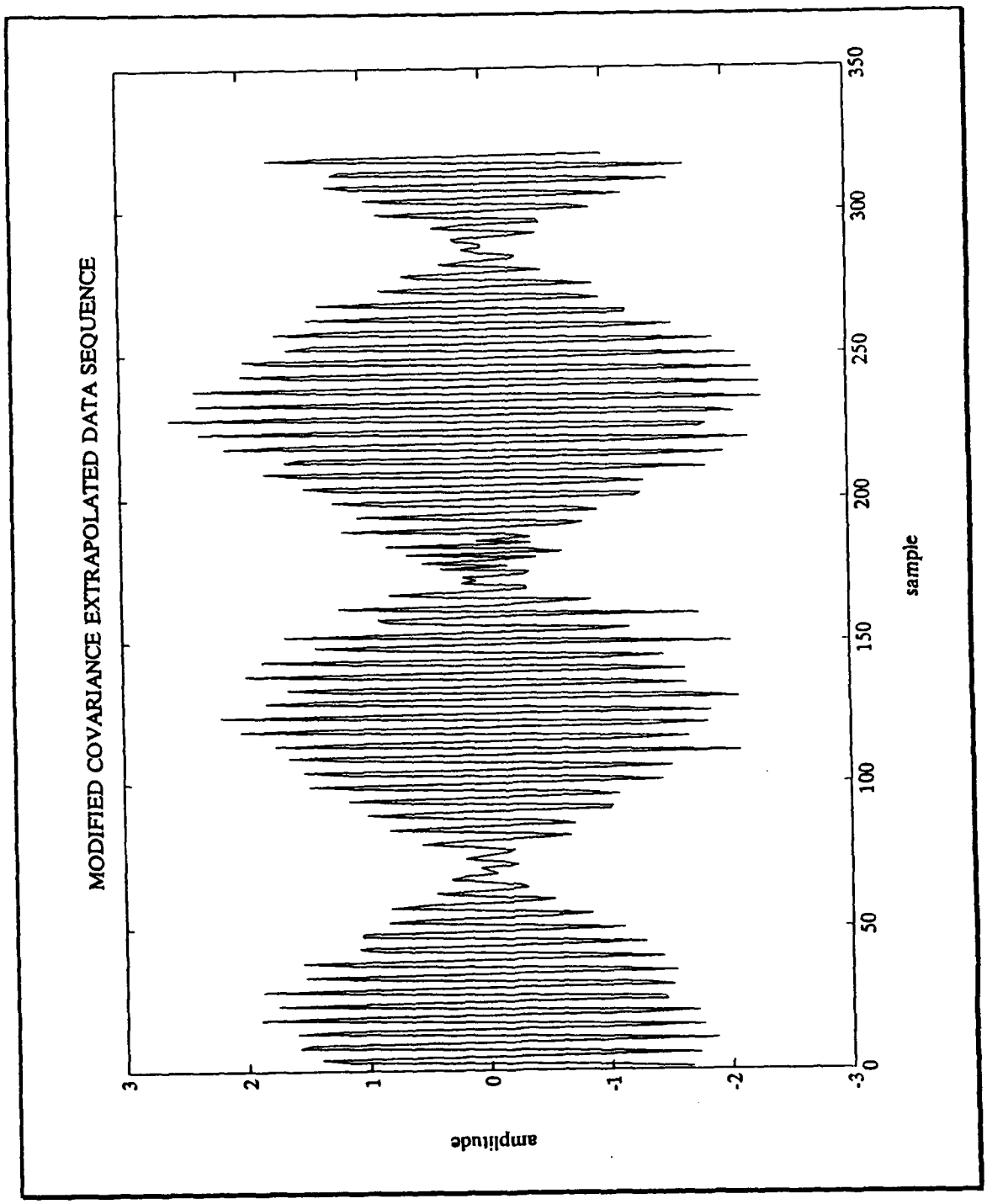

Figure 3. Extrapolated KM Data, Model Order = 16

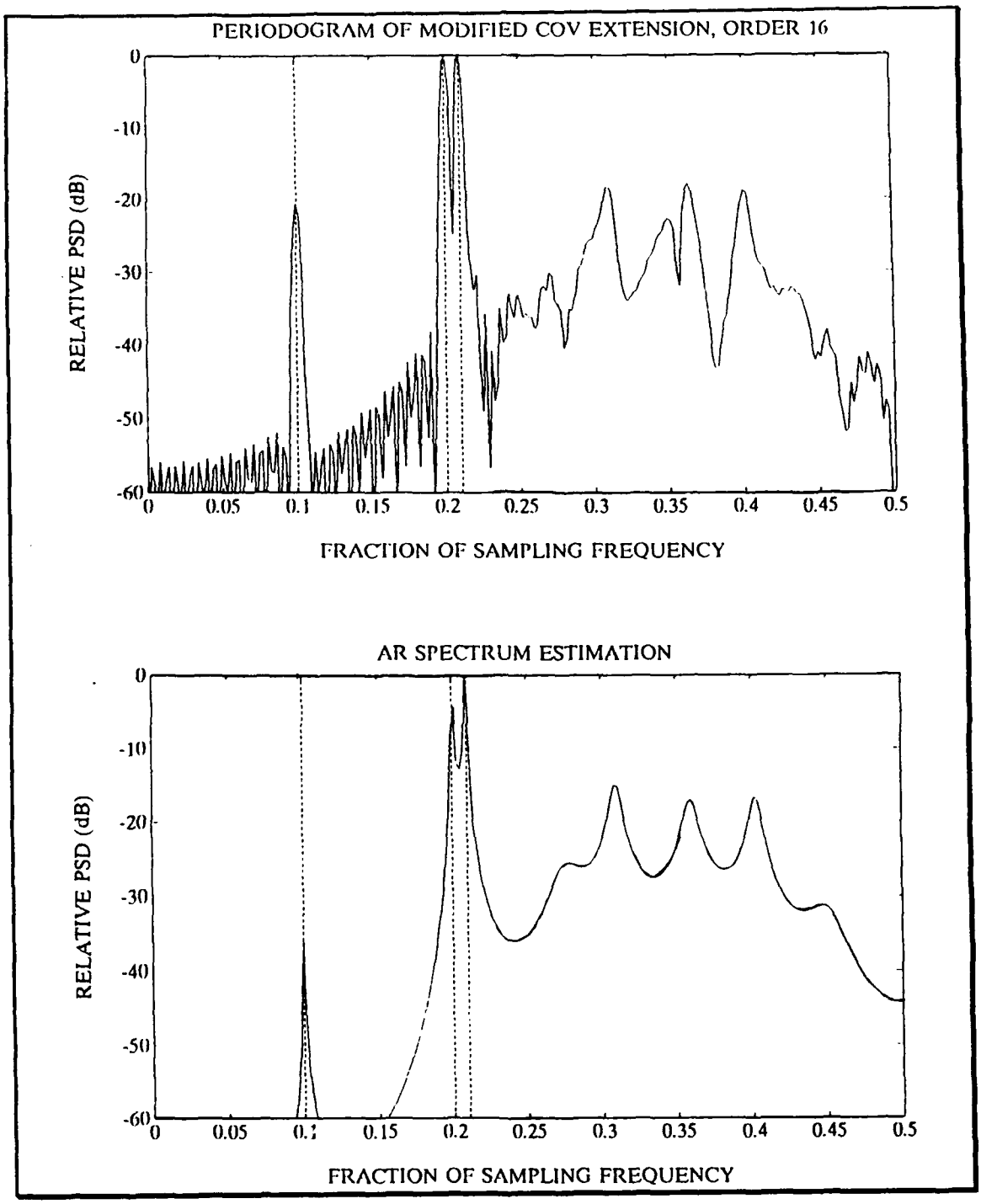

Figure 4. Periodogram of Extrapolated KM Data (320 points) and Respective AR Spectrum Estimation

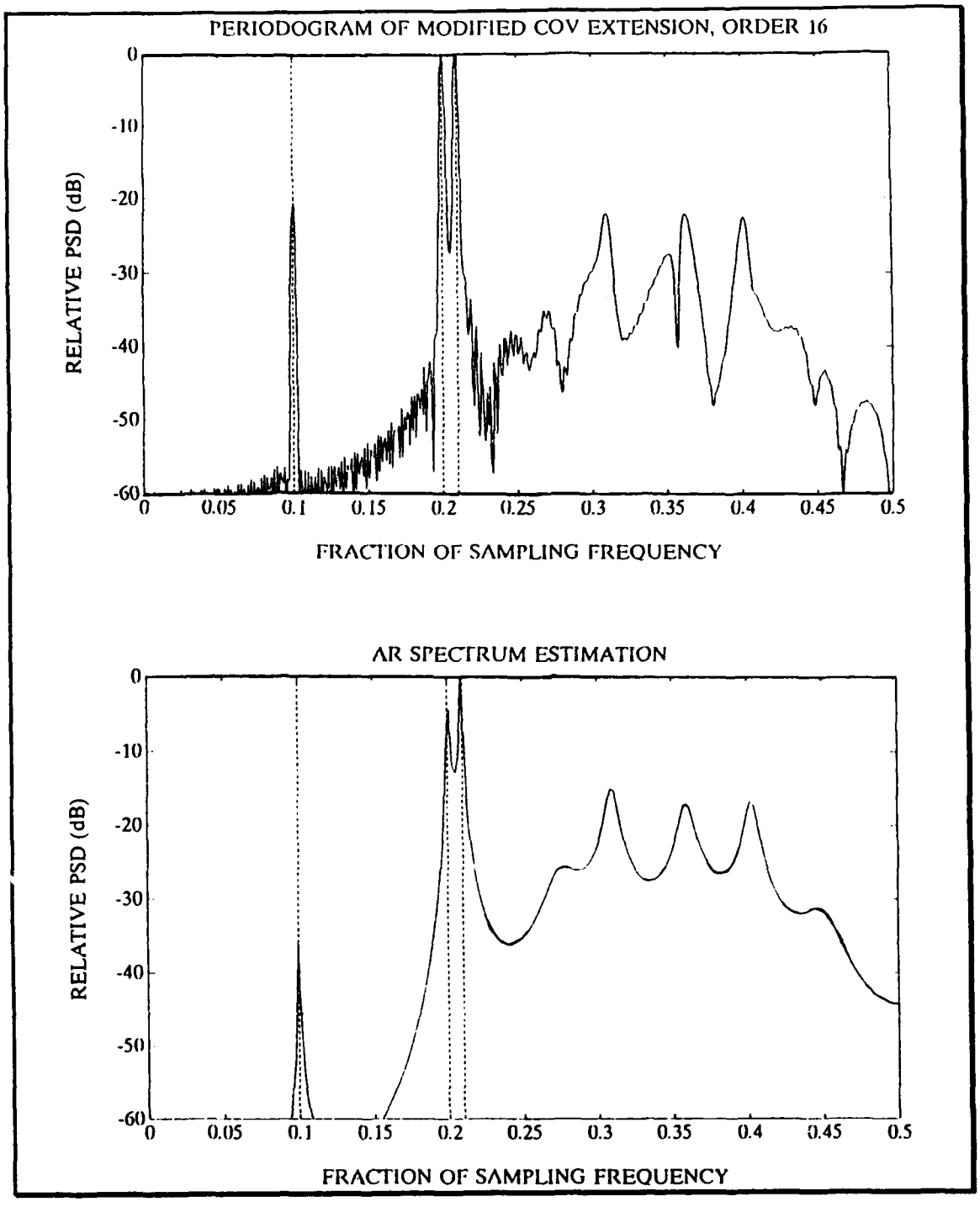

Figure 5. Periodogram of Extrapolated KM Data (576 points) and Respective AR Spectrum Estimation

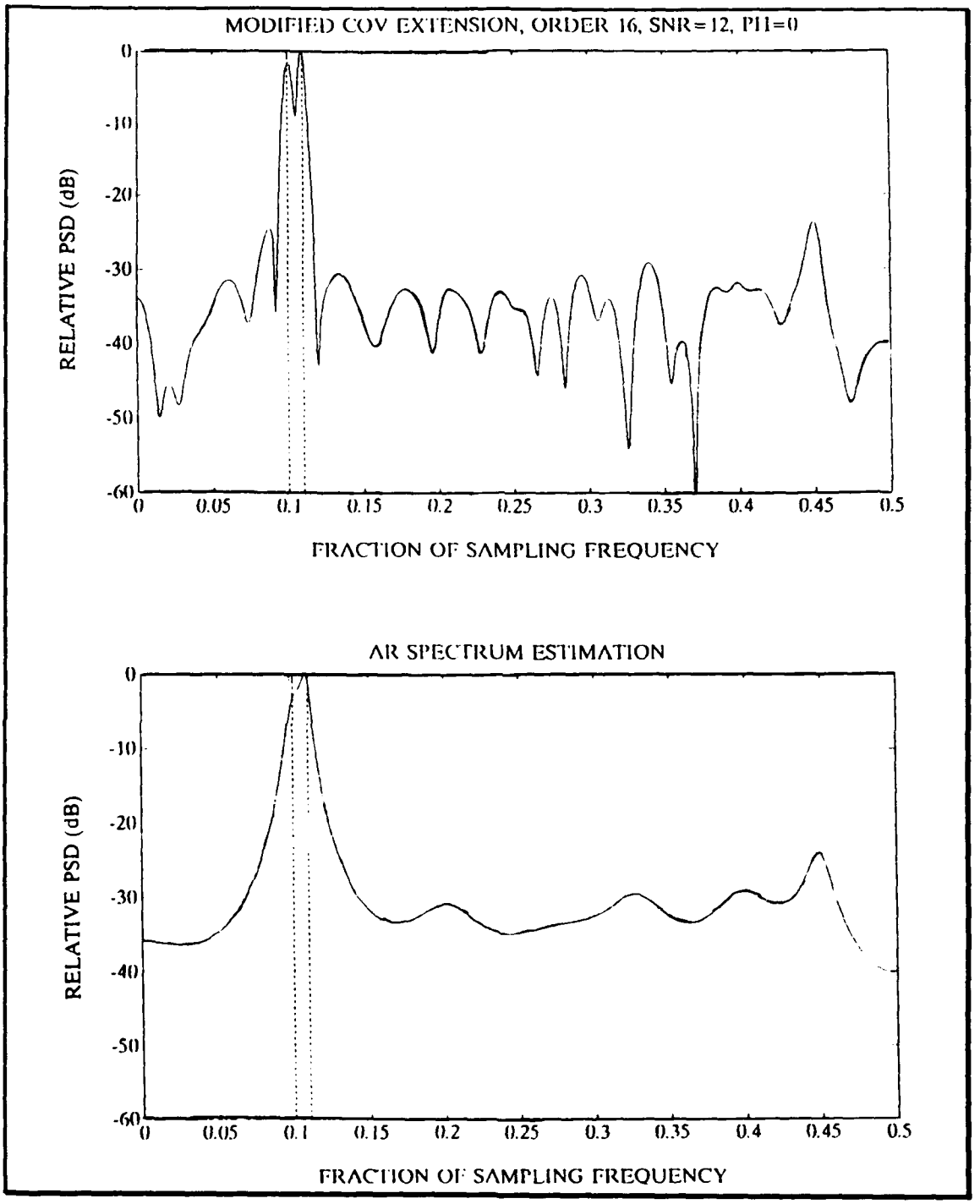

Figure 6. Periodogram of Extrapolated Sequence and Respective AR Spectrum Estimation of two Sine Waves (0° Initial Phase Angle)

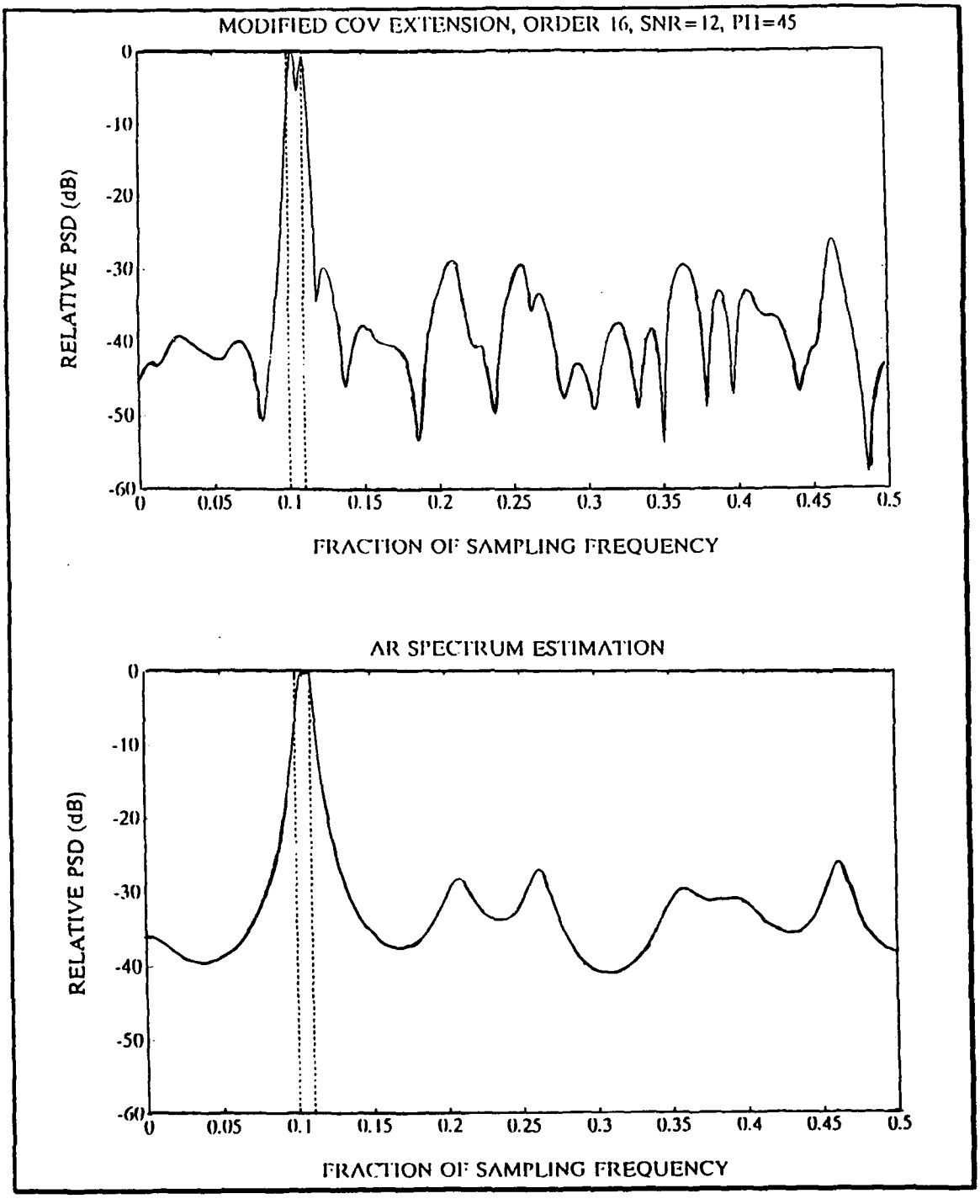

Figure 7. Periodogram of Extrapolated Sequence and Respective AR Spectrum Estimation of two Sine Waves (45° Initial Phase Angle)

When the initial phase angle between the two sine waves is changed to 90 degrees, in essence a signal consisting of a sine wave and a cosine wave with frequencies that are 10 Hz apart, neither method is able to distinguish the two frequencies.

2. Modified Forward-Backward Method

The modified forward-backward method is a two-step procedure for modifying the FBLP method [Ref. 3:p. 347]. First, we compute the P eigenvalues of the correlation matrix given in Eq. (2.12). To do this, we use the singular value decomposition (SVD) which is applied to the 2(N-P)-by-P data matrix A given by

$$A^H = \begin{bmatrix} x[P] & x[P+1] & \cdots & x[N-1] & x^*[2] & x^*[3] & \cdots & x^*[N-P+1] \\ x[P-1] & x[P] & \cdots & x[N-2] & x^*[3] & x^*[4] & \cdots & x^*[N-P+2] \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \cdots & \vdots \\ x[1] & x[2] & \cdots & x[N-P] & x^*[P+1] & x^*[P+2] & \cdots & x^*[N] \end{bmatrix} \tag{2.15}$$

where N is the length of the original data sequence. Note that the correlation matrix in Eq. (2.12) is equal to $A^H A$.

Once the P eigenvalues are computed, we divide the data space into two parts: (i) the signal subspace spanned by the eigenvectors associated with the K largest eigenvalues, and (ii) the noise subspace spanned by the remaining (P-K) eigenvectors. The notion used is that the K principal eigenvectors of the correlation matrix are due to the signal components and are less sensitive to perturbations caused by noise than the remaining eigenvectors. In the case of noiseless data consisting of K sinusoids, the correlation matrix has K nonzero eigenvalues and (P-K) zero eigenvalues. Thus in a practical case of noisy data, by keeping the K largest eigenvalues of the correlation matrix and ignoring the rest, we are in effect attempting to increase the SNR by approaching the noiseless case.

Let $\upsilon_1, \upsilon_2, ..., \upsilon_K$ denote the K largest eigenvalues and $\mathbf{e}_1, \mathbf{e}_2, ..., \mathbf{e}_K$ denote their associated eigenvectors, respectively. We use these eigenvalues and eigenvectors to compute the following P-by-1 estimate of the prediction coefficients:

$$\mathbf{a} = \sum_{i=1}^{K} \frac{\mathbf{e}_i}{\upsilon_i} \left(\mathbf{e}_i^H A^H \mathbf{b} \right) \tag{2.16}$$

where

$$\mathbf{b} = [x[P+1], x[P+2], \ldots, x[N], x^*[1], x^*[2], \ldots, x^*[N-P]]. \tag{2.17}$$

For the modified FBLP (MFBLP) method, we have removed the undesirable effects of the noise subspace eigenvectors, therefore we may increase the predictor order P. Tufts and Kumaresan [Ref. 7:pp. 975-989] have experimentally determined the value of the predictor order to be

$$P \cong 3N / 4 \tag{2.17}$$

where N is the original, nonextrapolated data length.

Since the computer simulation uses 64 data points, a model order of 48 is selected.

a. Modified FBLP Results

Many test runs are made on the signal consisting of 2 sine waves corrupted by white noise. Sinusoidal frequencies, phases, and SNRs are changed for comparison purposes. Table 1 specifies the parameters for each run and the corresponding figure number. In each test case, the data sequence is extrapolated to a data length of 320 points using linear prediction with the prediction coefficients obtained from the MFBLP method. Figures 8 through 24 show the results.

TABLE 1. PARAMETERS FOR TEST SEQUENCES

CASE #	SINUSOIDAL FREQUENCIES (HZ)	SNR (DB)	RELATIVE INITIAL PHASE	FIGURE #
1	100 and 107	12	0°	8
2	100 and 107	9	0°	9
3	100 and 107	6	0°	10
4	100 and 107	12	45°	11
5	100 and 107	9	45°	12
6	100 and 107	6	45°	13
7	100 and 107	3	45°	14
8	100 and 107	12	90°	15
9	100 and 107	9	90°	16
10	100 and 107	6	90°	17
11	100 and 107	3	90°	18
12	100 and 105	12	0°	19
13	100 and 105	12	45°	20
14	100 and 105	18	45°	21
15	100 and 105	12	90°	22
16	100 and 105	18	90°	23
17	100 and 102	37	0°	24

In all test cases, the periodogram of the MFBLP extrapolated sequence performs as well or better than the AR spectrum estimation. Figure 8 shows that even though both methods correctly estimate the frequencies, the periodogram of the MFBLP extrapolated sequence sometimes has better defined spectral peaks. In case 10, depicted in Figure 17, the periodogram of the MFBLP extrapolated sequence correctly distinguishes the two frequencies while the AR spectrum estimation is unable to do so. Even in test case 17 where the two sine waves are only 2 Hz apart, the periodogram of the MFBLP extrapolated sequence has well defined peaks, as shown in Figure 25, while the AR spectrum estimation just barely distinguishes the two frequencies.

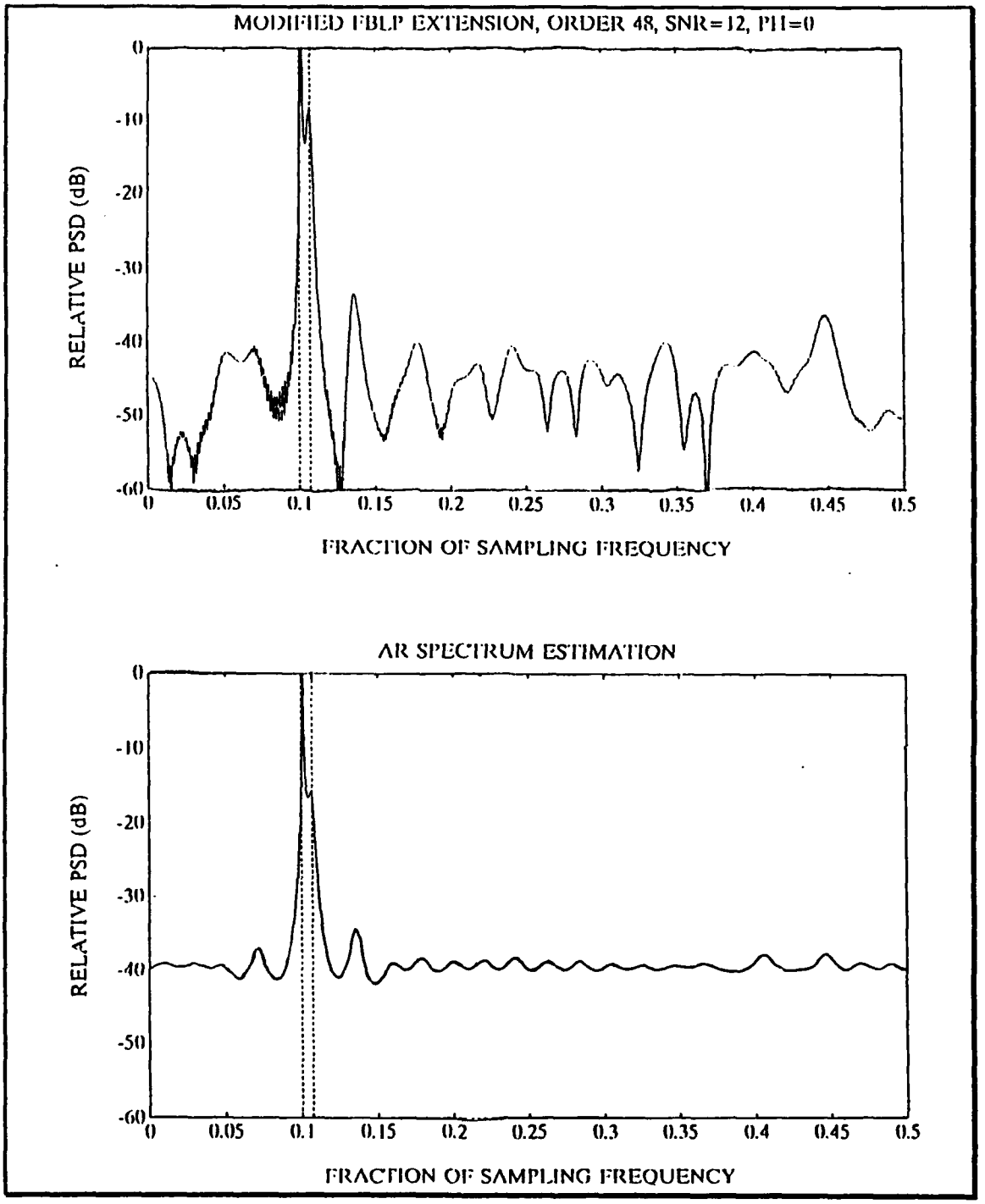

Figure 8. Case 1 Spectrum Estimates

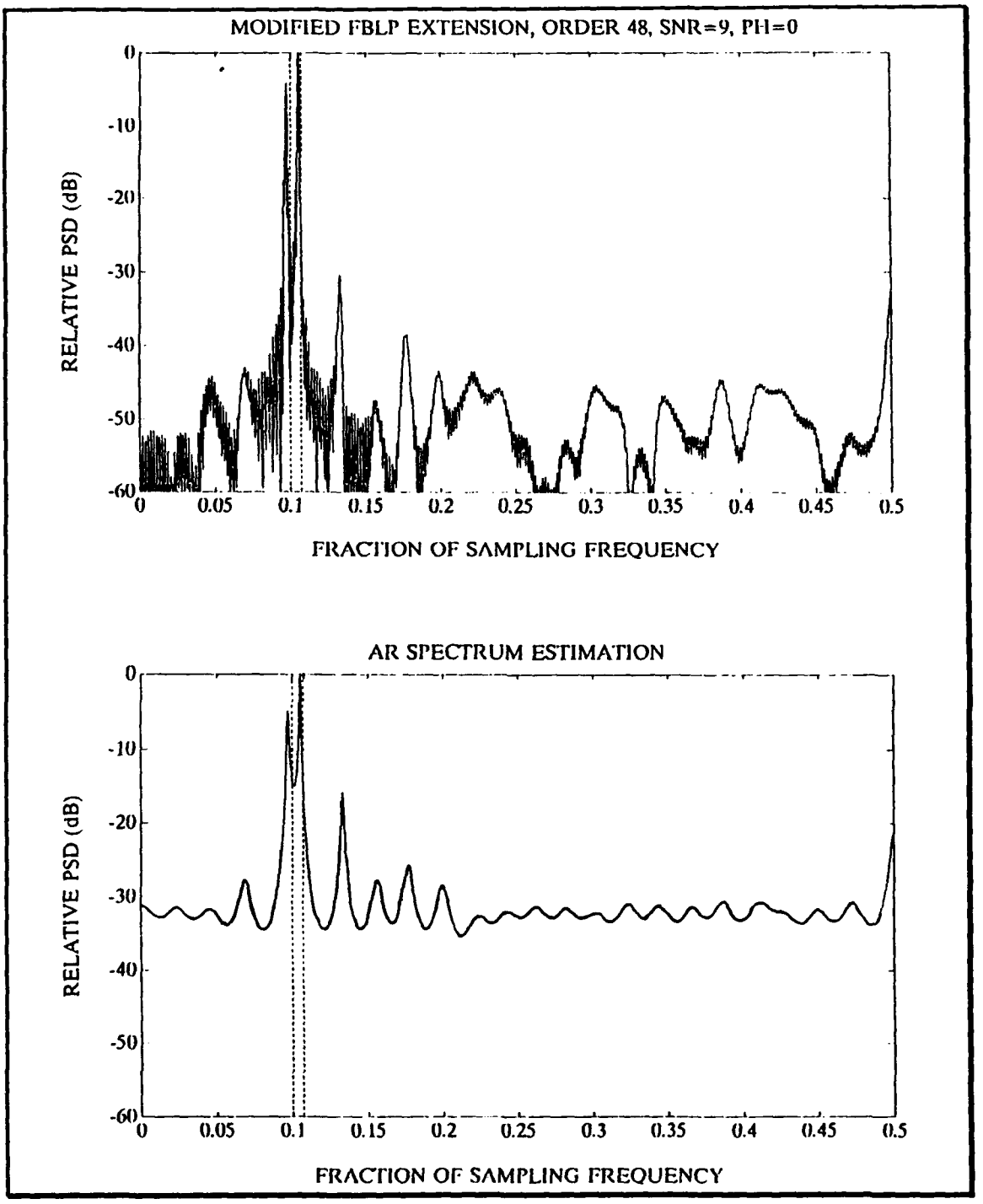

Figure 9. Case 2 Spectrum Estimates

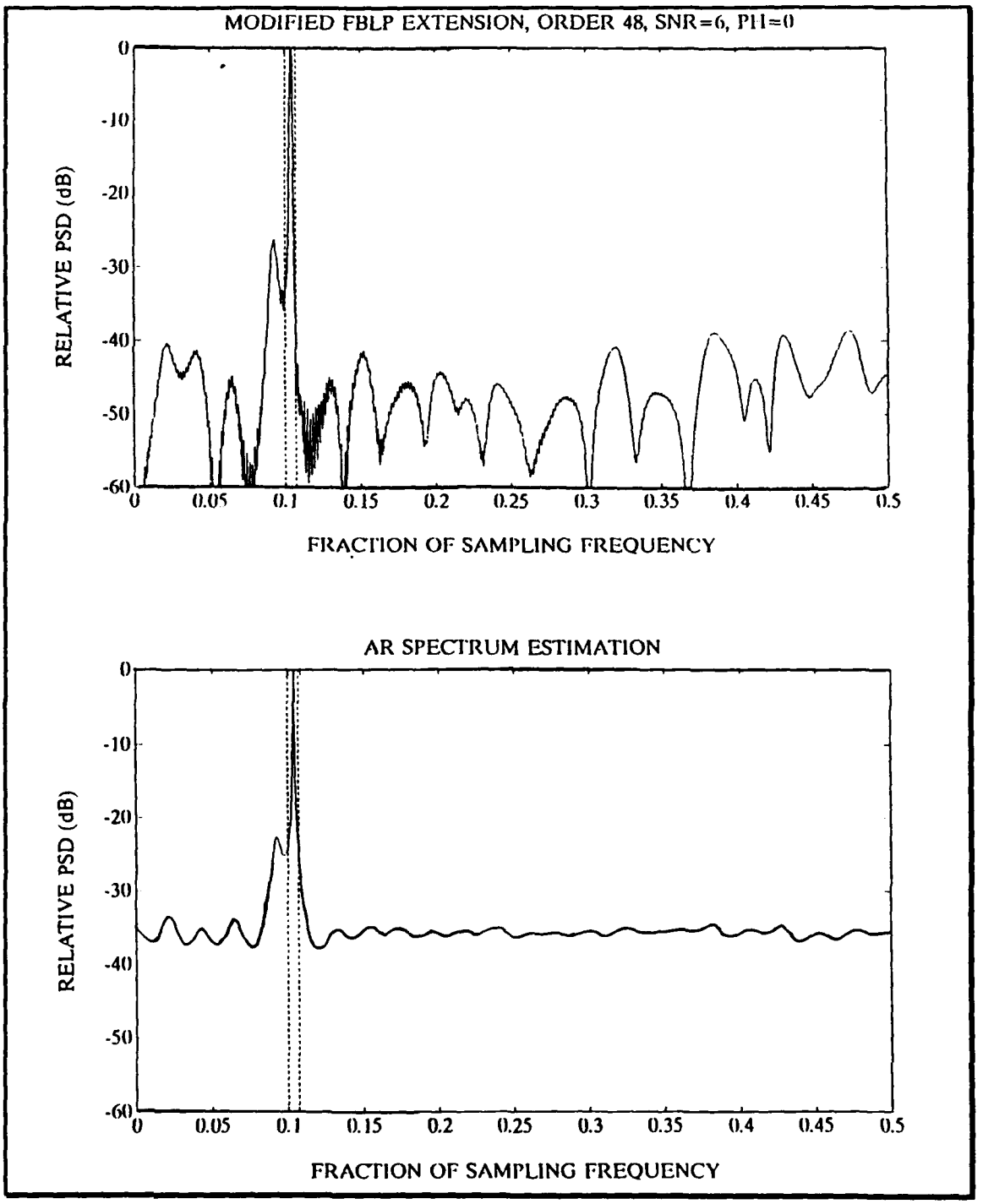

Figure 10. Case 3 Spectrum Estimates

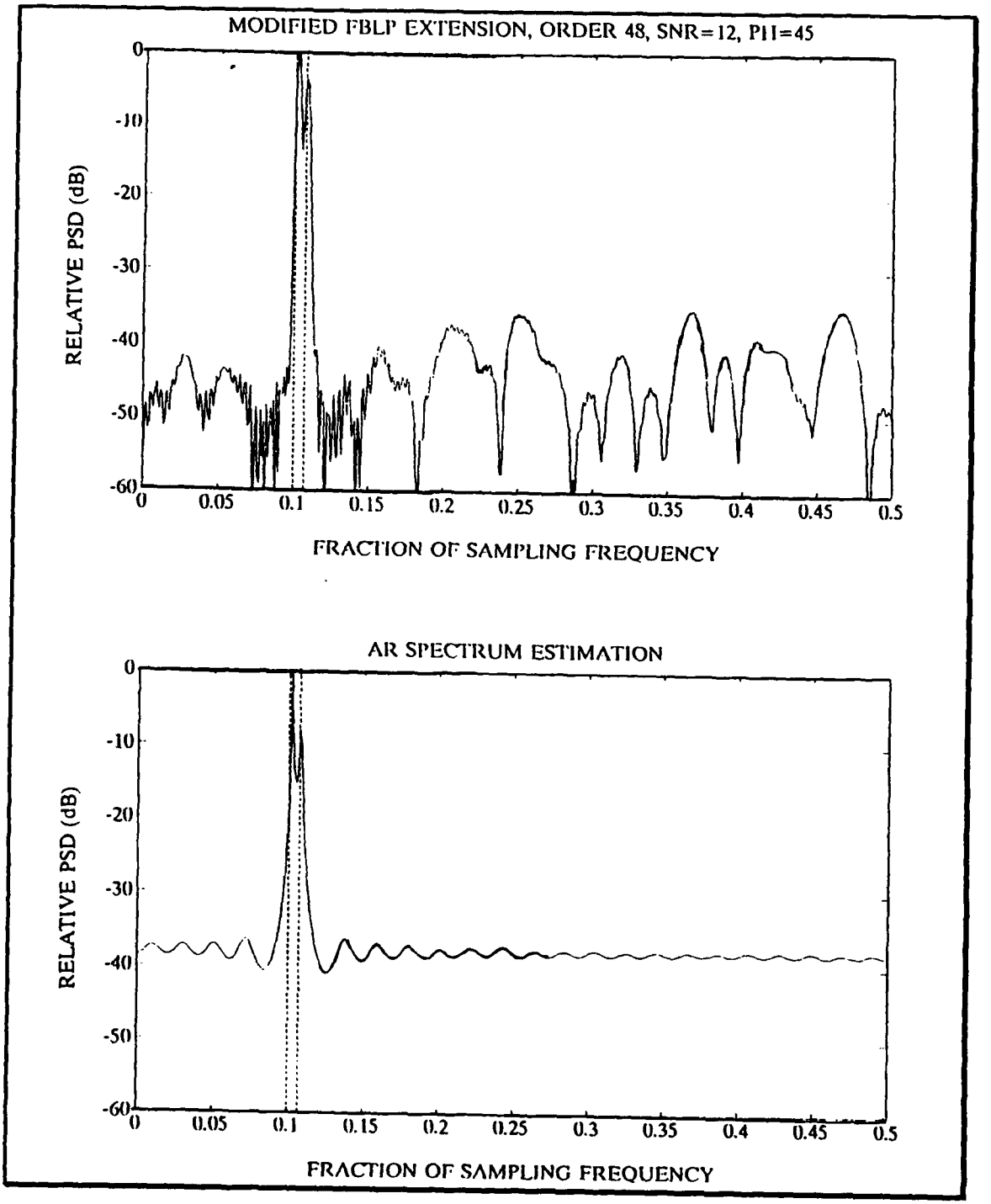

Figure 11. Case 4 Spectrum Estimates

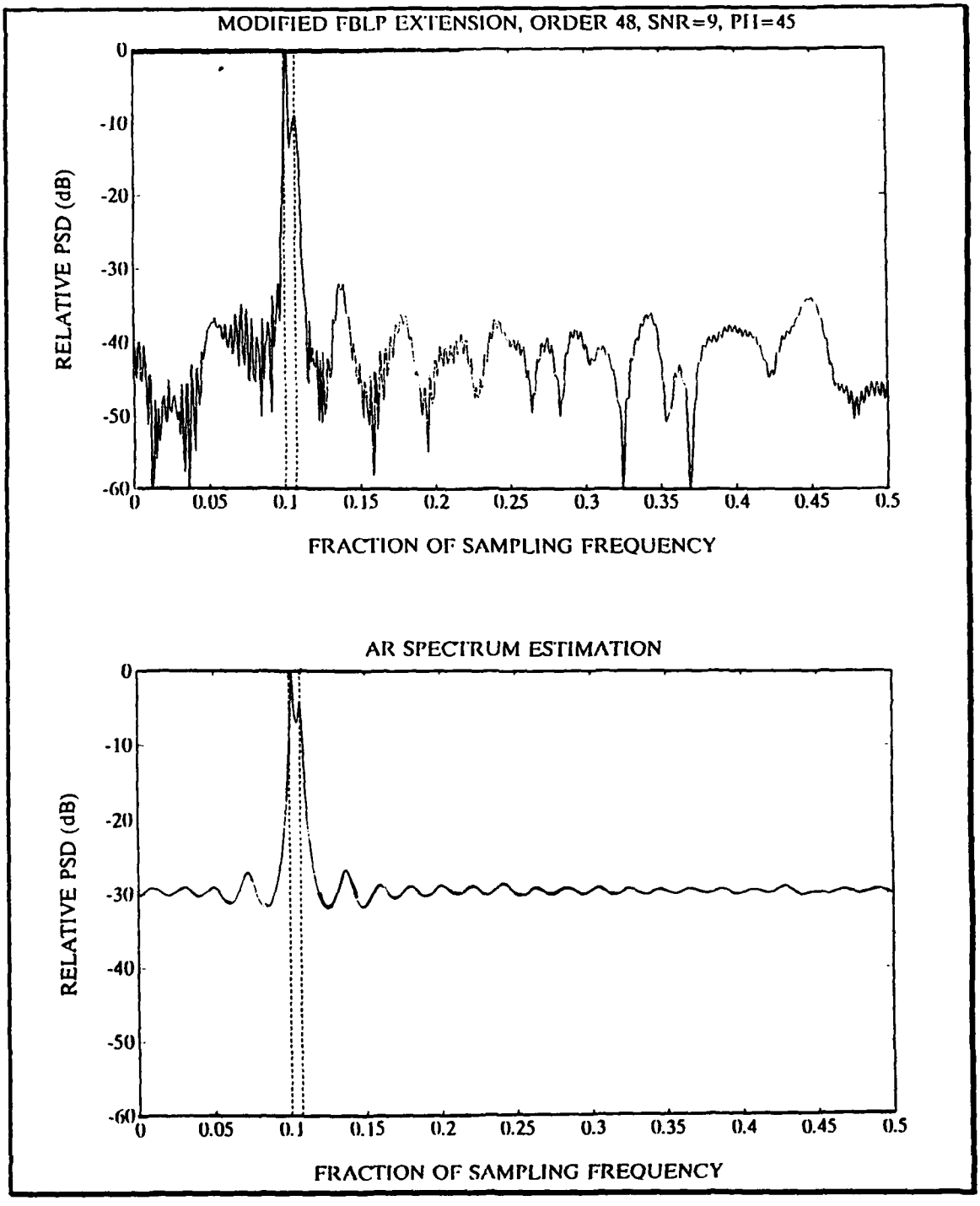

Figure 12. Case 5 Spectrum Estimates

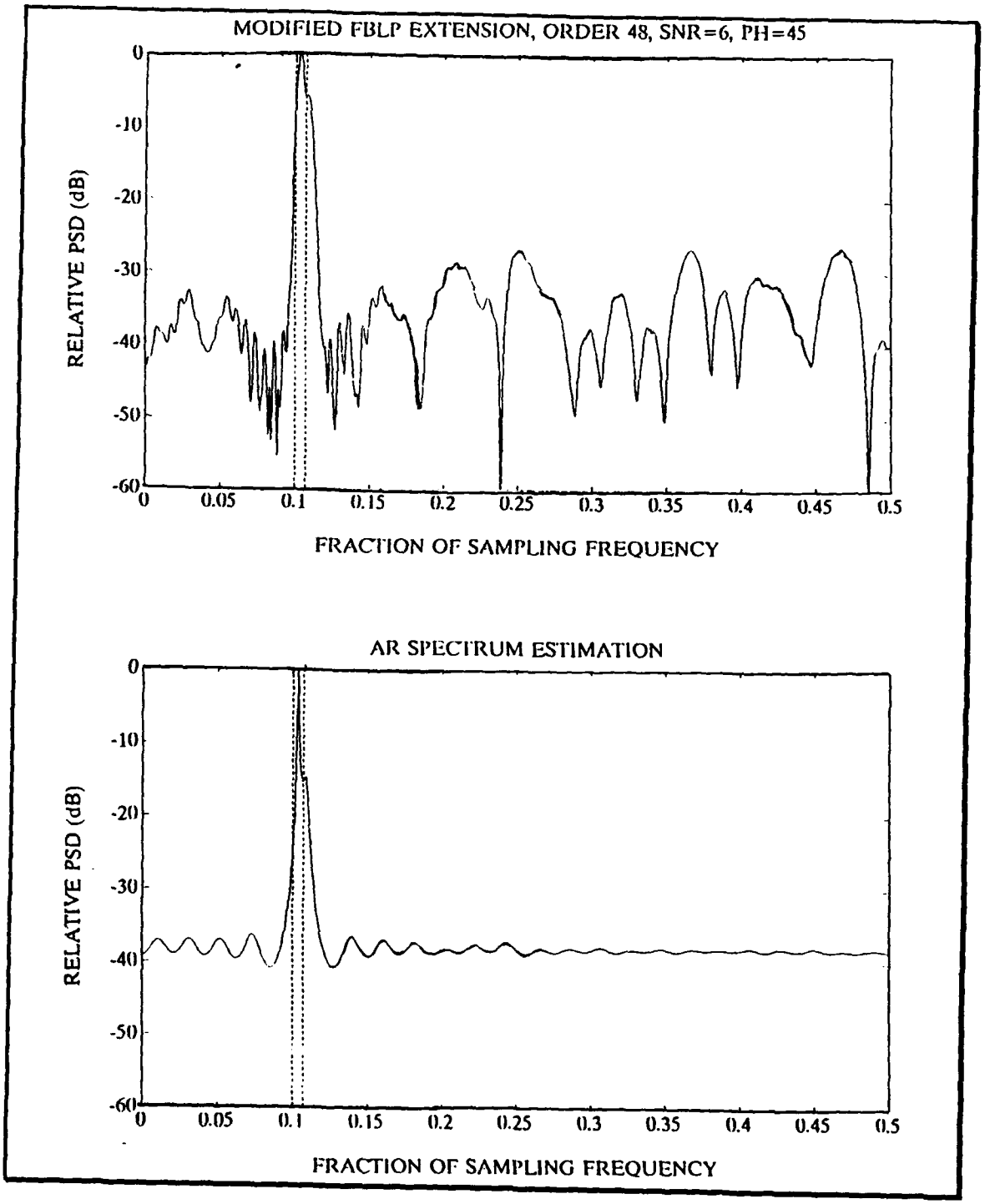

Figure 13. Case 6 Spectrum Estimates

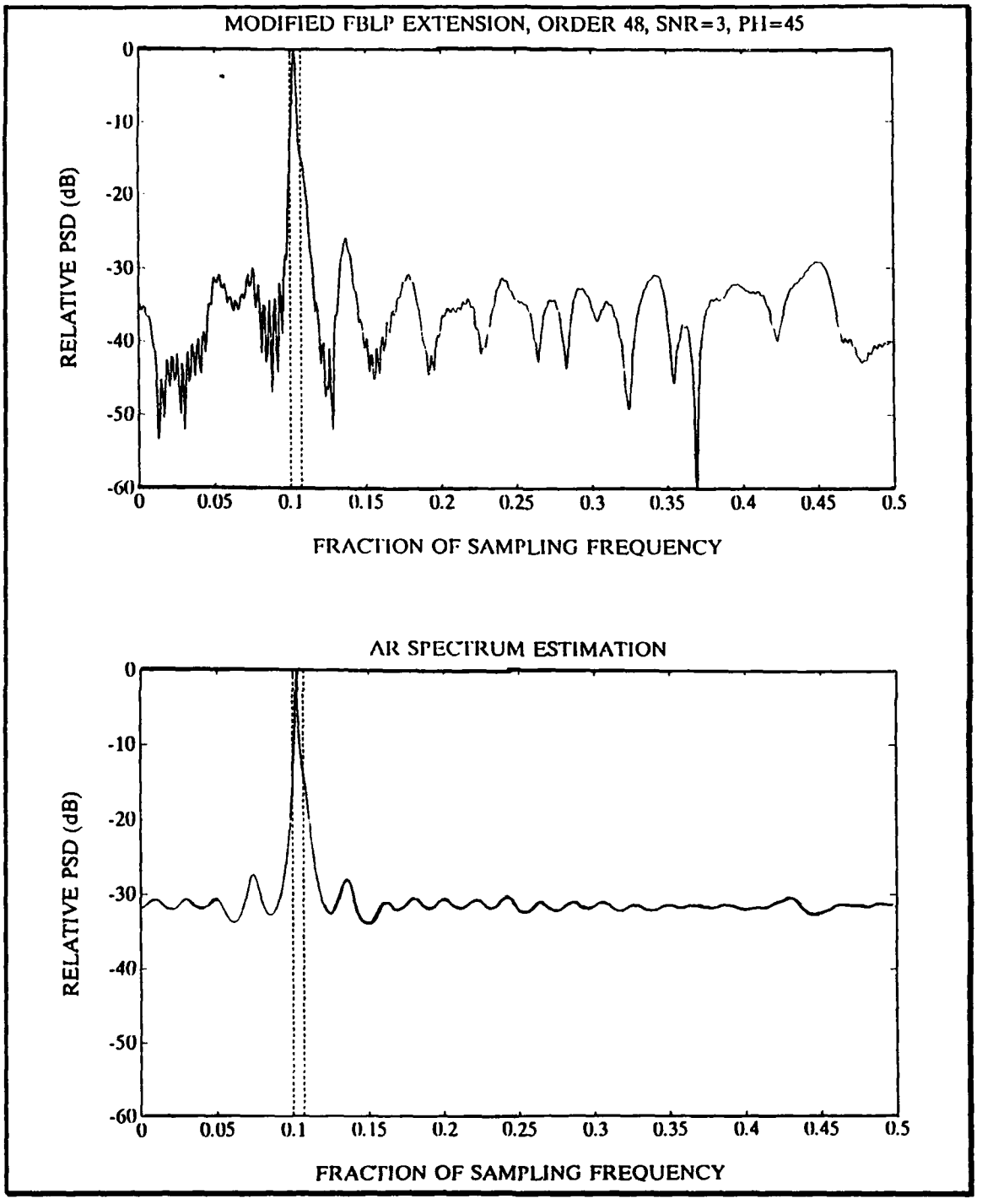

Figure 14. Case 7 Spectrum Estimates

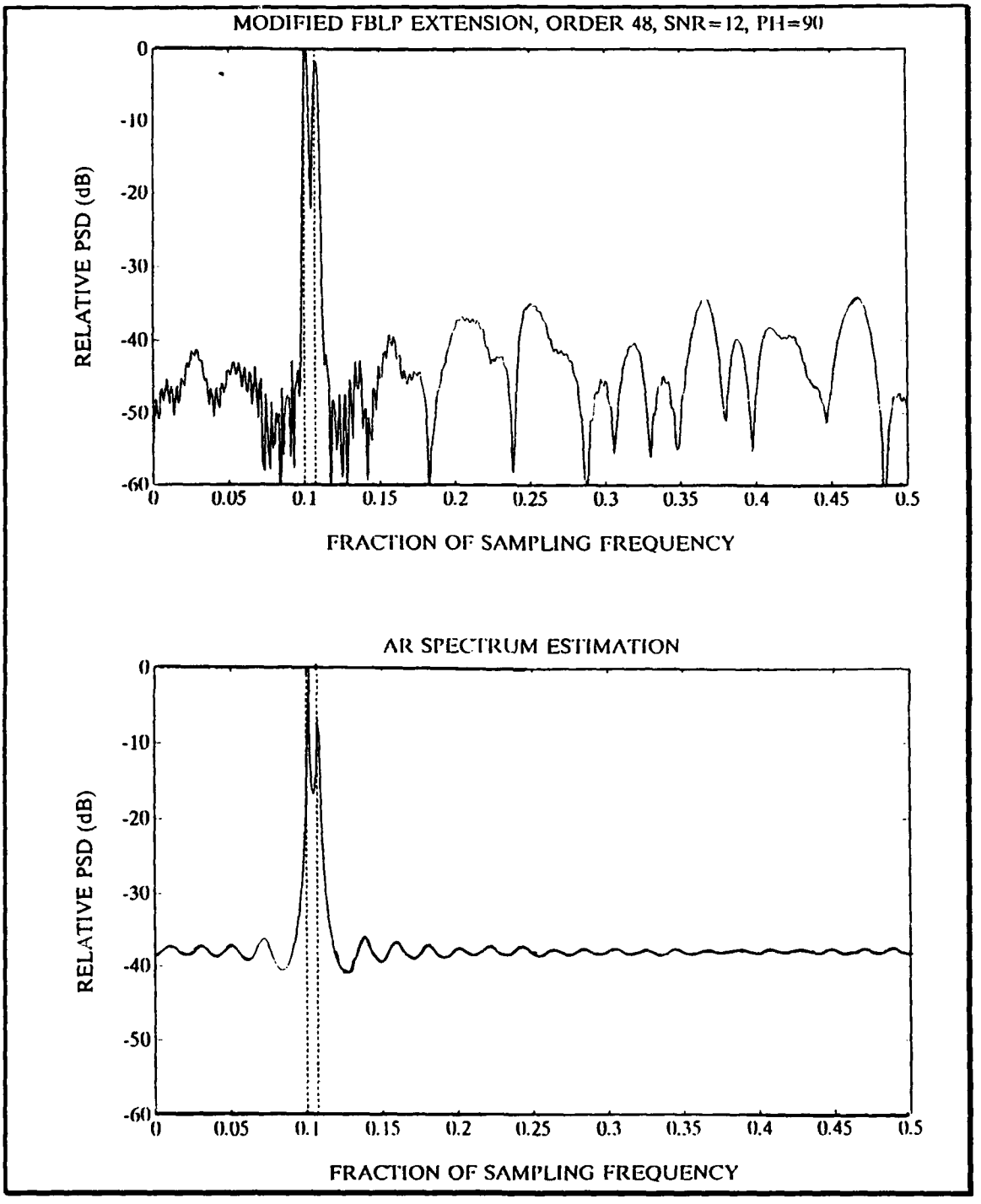

Figure 15. Case 8 Spectrum Estimates

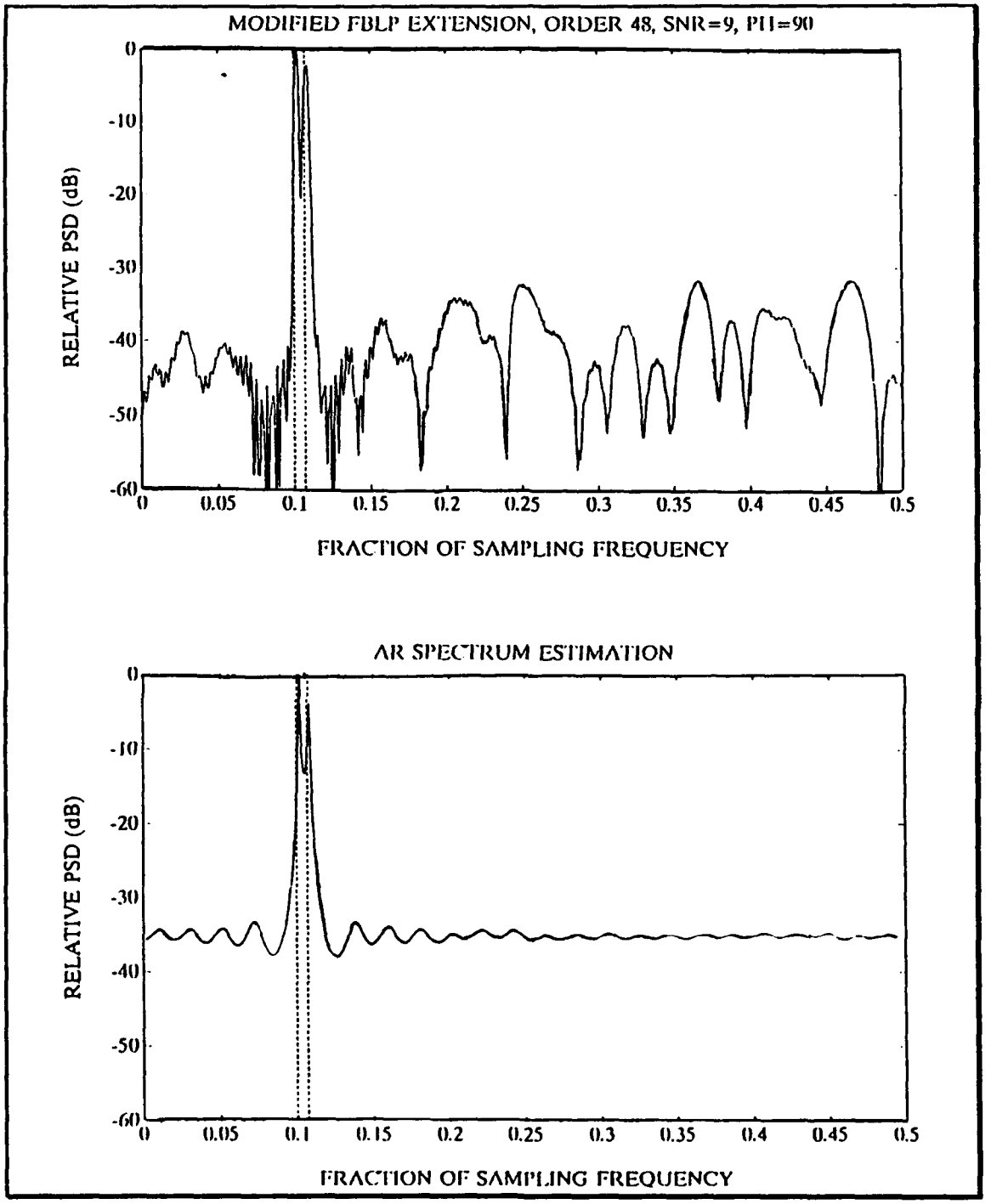

Figure 16. Case 9 Spectrum Estimates

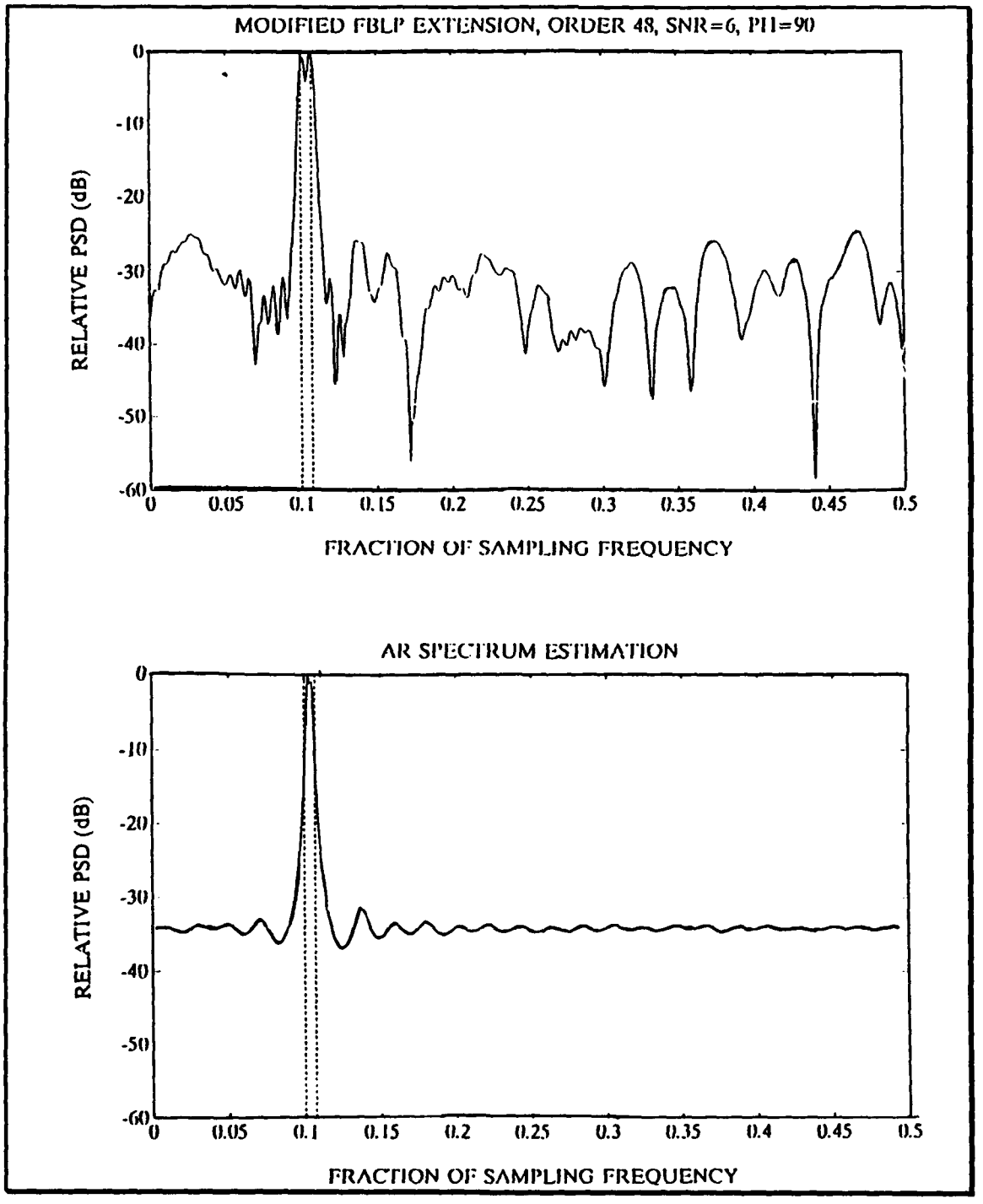

Figure 17. Case 10 Spectrum Estimates

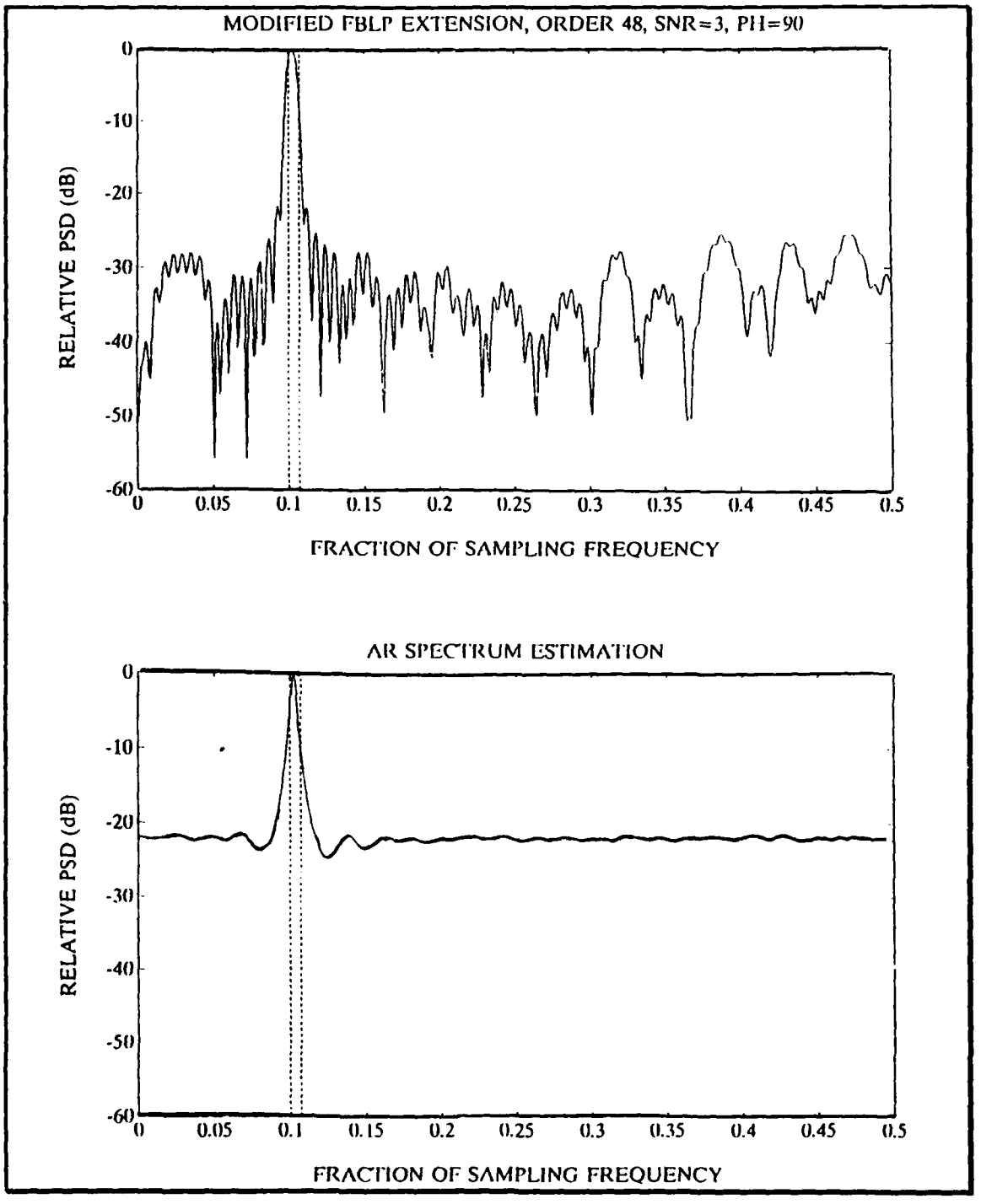

Figure 18. Case 11 Spectrum Estimates

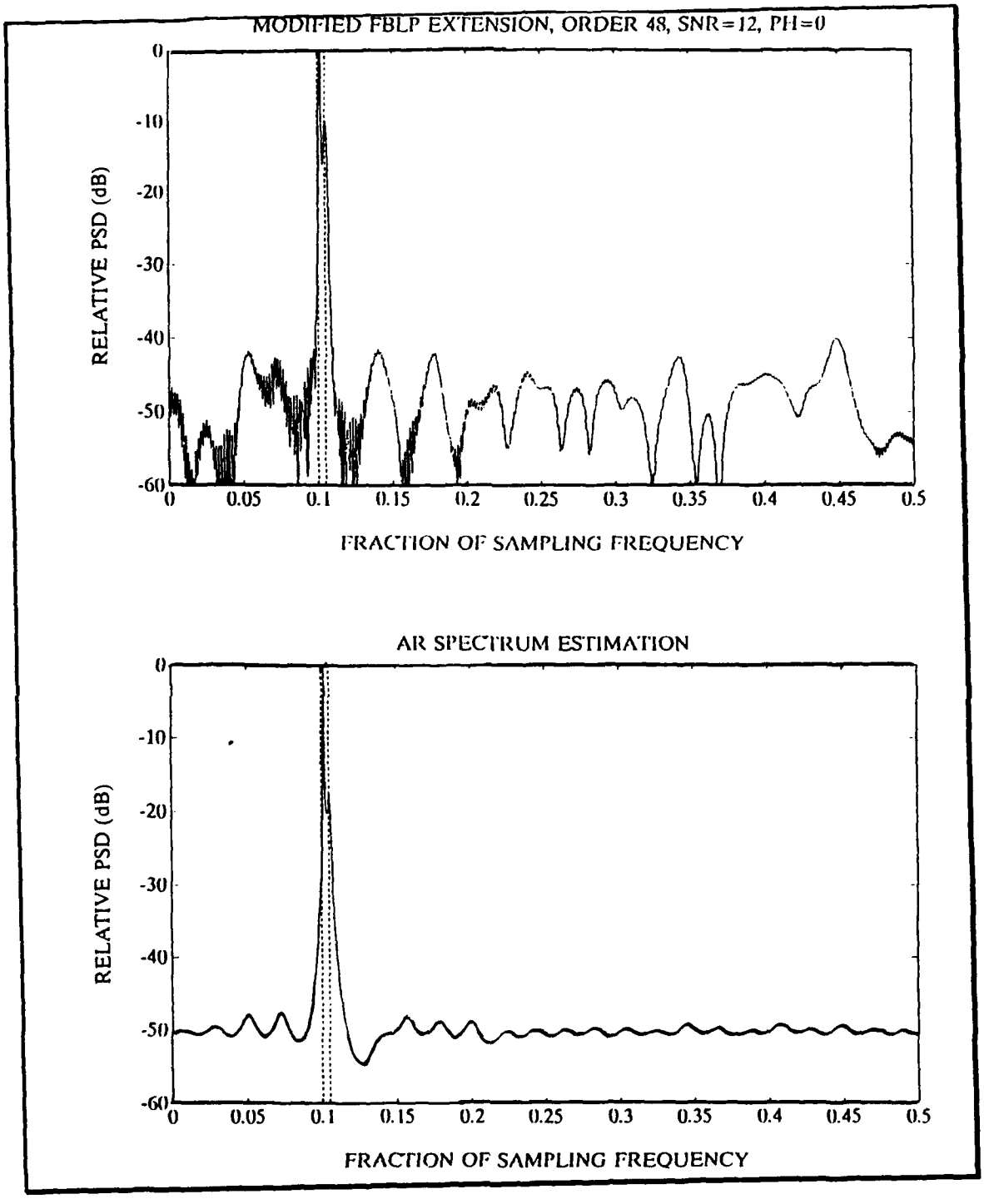

Figure 19. Case 12 Spectrum Estimates

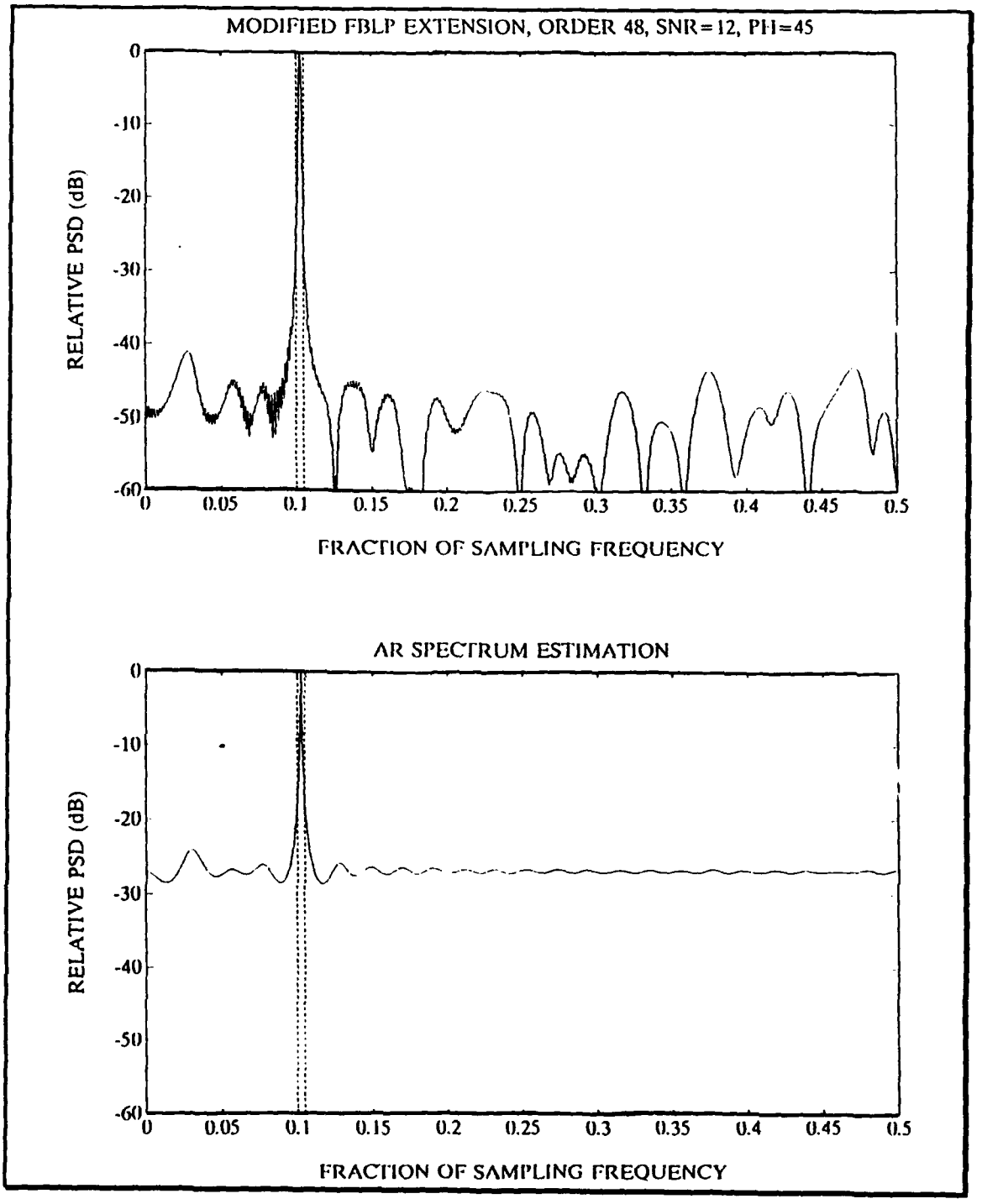

Figure 20. Case 13 Spectrum Estimates

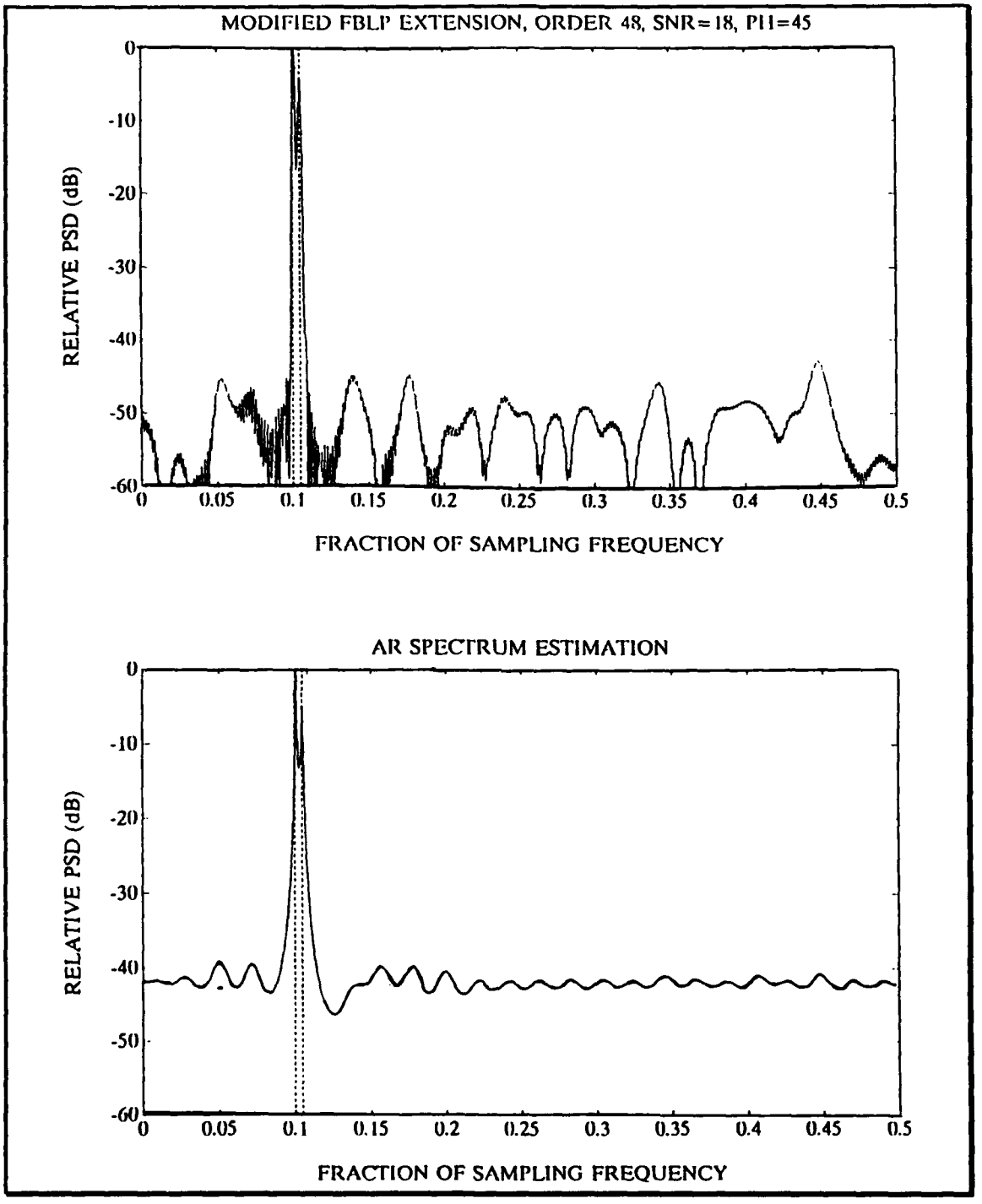

Figure 21. Case 14 Spectrum Estimates

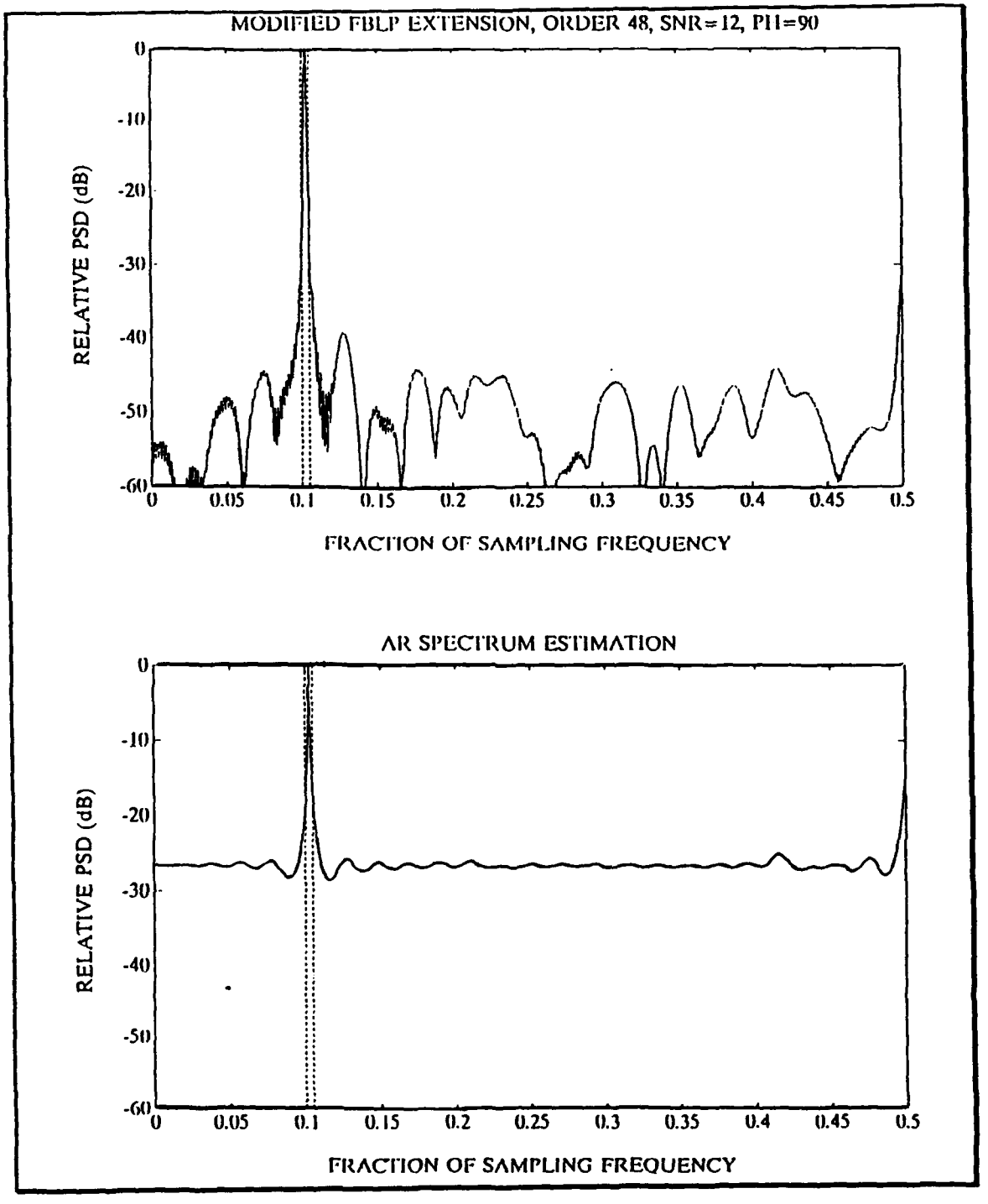

Figure 22. Case 15 Spectrum Estimates

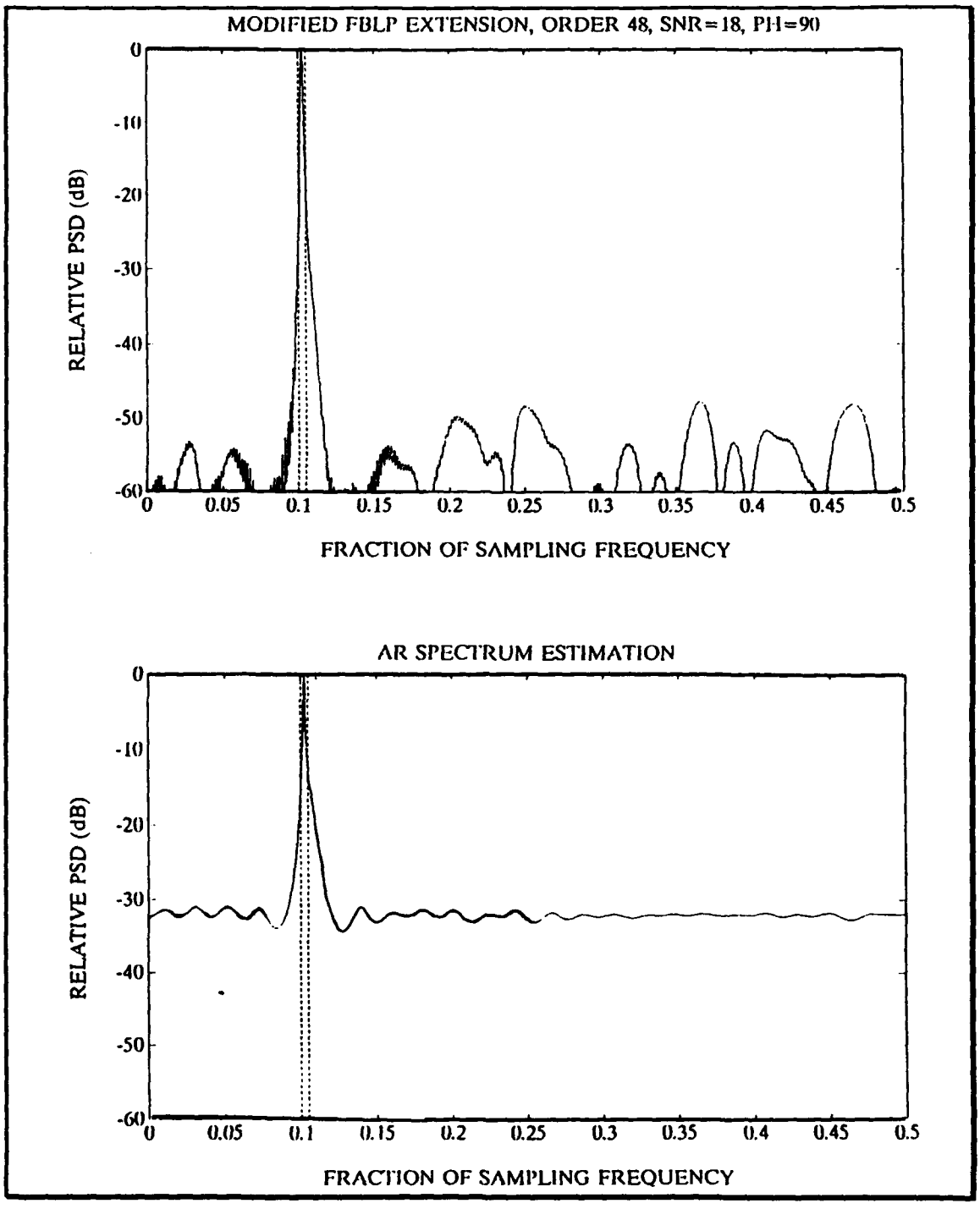

Figure 23. Case 16 Spectrum Estimates

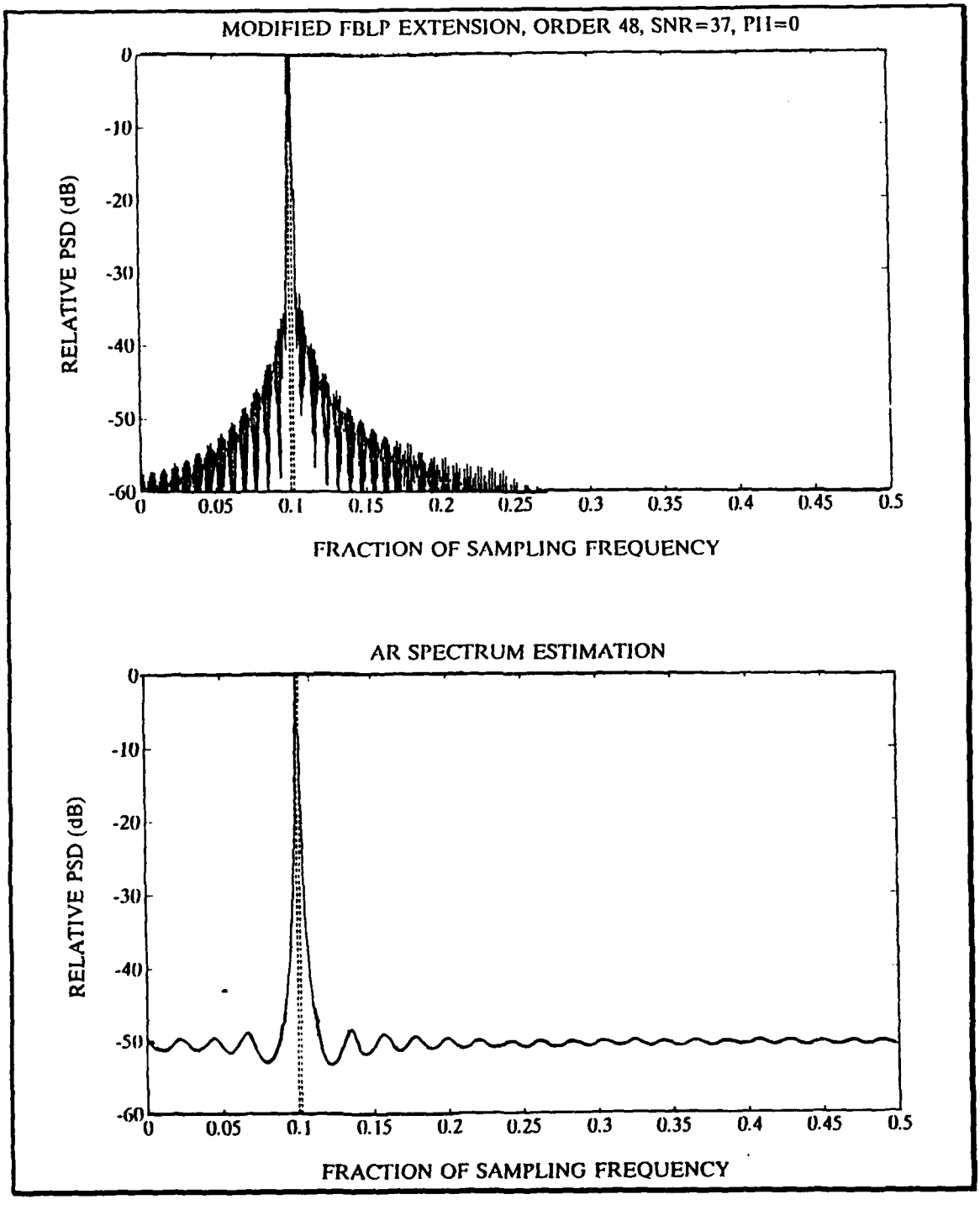

Figure 24. Case 17 Spectrum Estimates

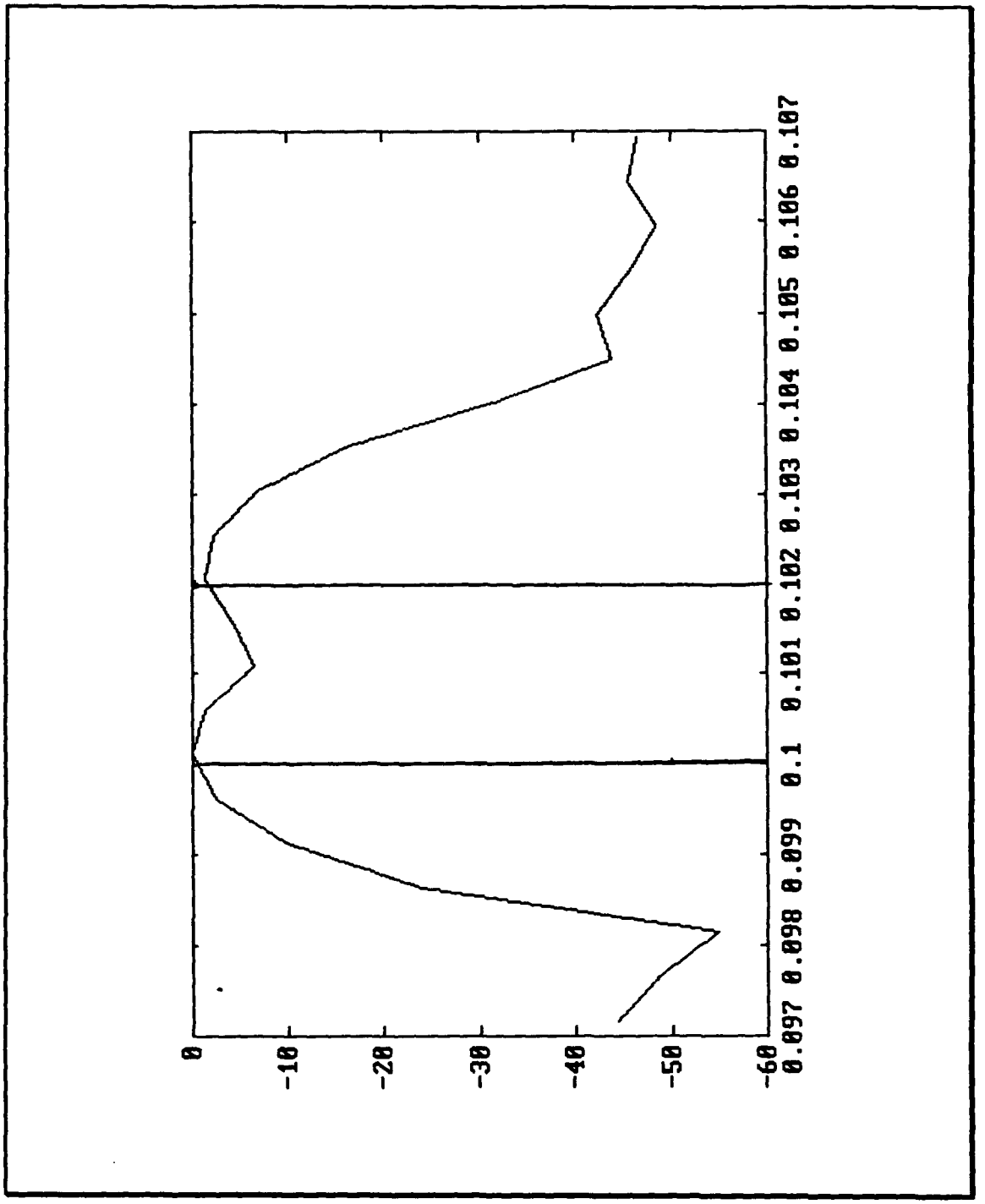

Figure 25. Periodogram of Extrapolated Data Sequence for Case 17

When the MFBLP extrapolation technique is tested on the KM sequence, the periodogram of the extrapolated sequence is considerably better than the AR spectral based estimation. In general, the modified FBLP method utilizes prior knowledge of the number of sinusoids present in the signal. When the K largest eigenvalues are chosen, we are assuming that there are K sinusoids present in the signal. If one of these K largest eigenvalues is associated with the noise, inaccurate estimates will be computed. This can be seen in Figure 26. Because the KM data is made up of 3 sinusoids, the 3 largest eigenvalues and corresponding eigenvectors are chosen. However in this case, one of these eigenvalues is associated with the noise so the signal at the 0.1 fractional frequency is not detected in the AR spectral estimation. However when the same prediction coefficients are used in the data extrapolation, the resulting periodogram is a more accurate spectral estimation than the AR spectral estimation. Obviously, the extrapolation approach is more forgiving if the exact number of sinusoids is unknown. In Figure 27 the four largest eigenvalues and corresponding eigenvectors were chosen, one more than the actual number of sinusoids present. Once again the AR spectral estimation is incorrect, picking up more noise, while the periodogram of the modified FBLP extrapolation gives a reasonable spectral estimate. The KM sequence is extrapolated to a length of 320 points before the periodograms are computed.

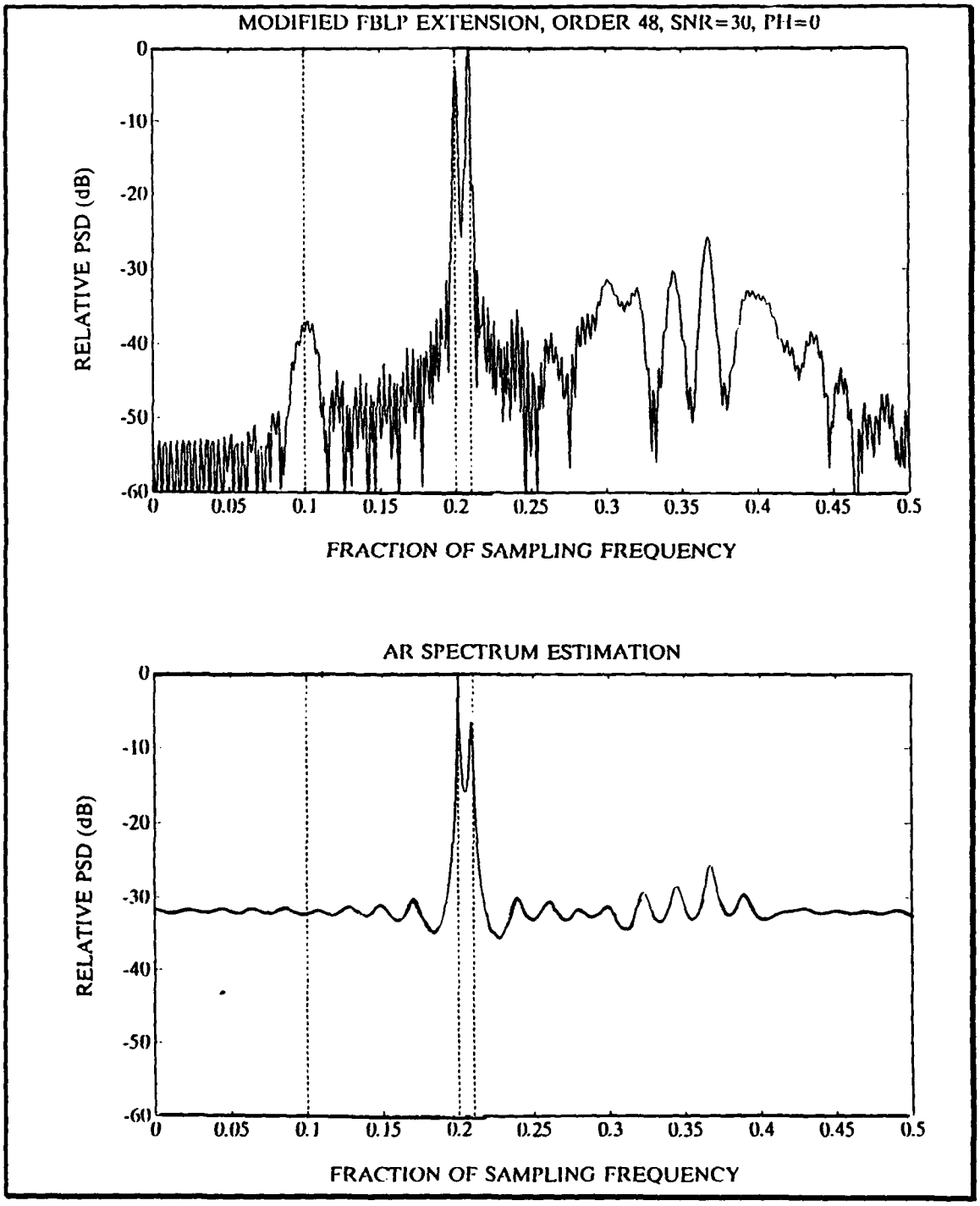

Figure 26. KM Spectrum Estimates Using the Three Dominant Eigenvectors

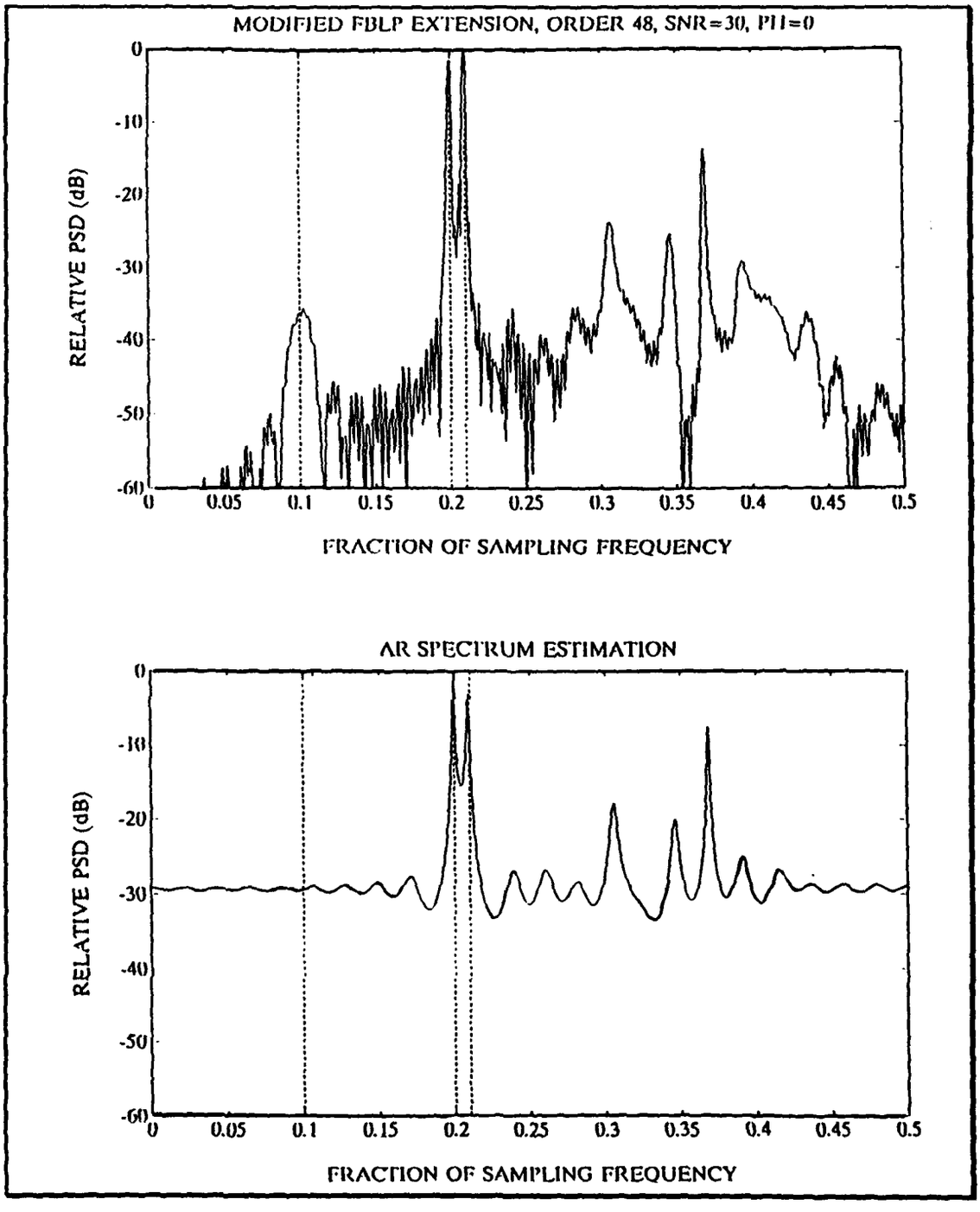

Figure 27. KM Spectrum Estimates Using the Four Dominant Eigenvectors

III. PRONY'S METHOD

A. DEFINITION

The second method of sequence extrapolation involves modeling the data as the impulse response of a linear time invariant filter. This approach is shown in Figure 28. Note that B(z) has q zeros and A(z) has p poles.

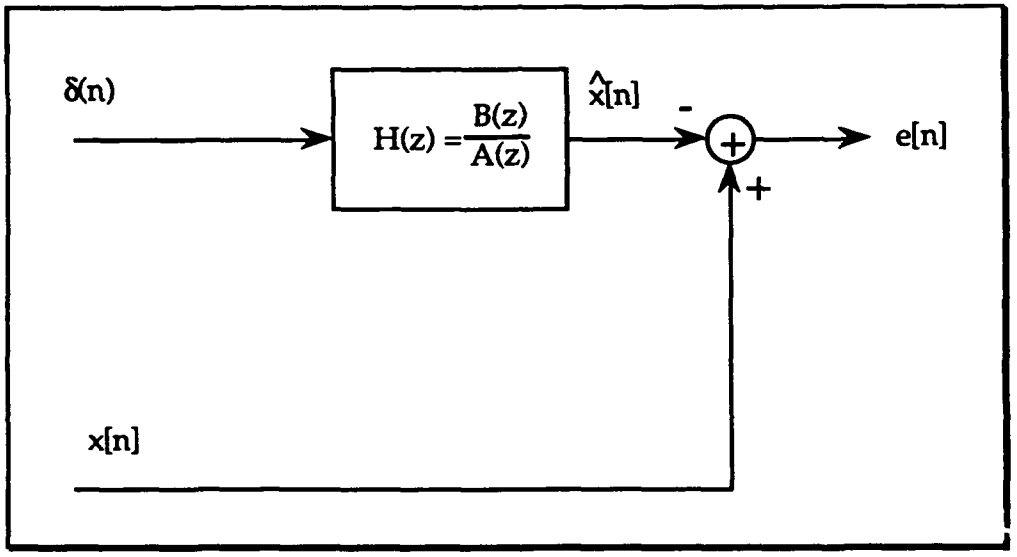

Figure 28. Impulse Response Estimation

We want to minimize the error e[n] in some sense. If we would minimize $\sum|e[n]|^2$, then this leads to non-linear equations and is very difficult to solve. Therefore we will use an indirect method to solve this problem. The equation for the filter is

$$\hat{x}[n]+a_1\hat{x}[n-1]+\ldots+a_p x[n-p]=b_0\delta(n)+b_1\delta(n-1)+\ldots+b_q\delta(n-q). \qquad (3.1)$$

The coefficients a_i and b_i are unknown in Eq. (3.1) and we want the output estimate $\hat{x}[n]$ as close to x[n] as possible.

Assume that we have N data points and that N ≥ p+q+1. The transfer function H(z) of the filter is

$$H(z) = \frac{B(z)}{A(z)} = \frac{b_0 + b_1 z^{-1} + \ldots + b_q z^{-q}}{1 + a_1 z^{-1} + \ldots a_p z^{-p}} \tag{3.2}$$

We can write Eq. (3.1) in matrix form as

$$\left[\begin{array}{ccccc} x[0] & 0 & 0 & \cdots & 0 \\ x[1] & x[0] & 0 & \cdots & 0 \\ \vdots & & & & \\ x[q] & x[q-1] & \cdots & \cdots & x[q-p] \\ \hline x[q+1] & x[q] & \cdots & \cdots & x[q+1-p] \\ \vdots & & & & \\ x[N-1] & x[N-2] & \cdots & \cdots & x[N-1-p] \end{array}\right] \left[\begin{array}{c} 1 \\ a_1 \\ a_2 \\ \vdots \\ a_p \end{array}\right] = \left[\begin{array}{c} b_0 \\ b_1 \\ b_2 \\ \vdots \\ b_q \\ \hline 0 \\ \vdots \\ 0 \end{array}\right] \tag{3.3}$$

or more compactly

$$\begin{bmatrix} X_B \\ X_A \end{bmatrix} \mathbf{a} = \begin{bmatrix} \mathbf{b} \\ \mathbf{0} \end{bmatrix} \tag{3.4}$$

where $\mathbf{0}^T = [0, 0, ..., 0]$.

For the case where N=p+q+1, we can write the lower set of equations as

$$X_A \mathbf{a} = \mathbf{0} \tag{3.5}$$

and solve directly for **a**. Once we know **a** we can solve the upper equations directly for **b** where

$$X_B \mathbf{a} = \mathbf{b}. \tag{3.6}$$

This method is known as Pade's approximation [Ref 1:p. 495].

For the more general case where N > p+q+1 we still have Eqs. (3.5) and (3.6), however now Eq. (3.5) is an overdetermined set. If we solve Eq. (3.5) in a least squares sense, then we can still find **b** from Eq. (3.6). Solution of this

least squares problem [Ref. 1:pp. 474-495] leads to the least squares normal equations

$$\left(X_A^{*T} X_A\right)\mathbf{a} = \begin{bmatrix} S \\ \mathbf{0} \end{bmatrix} \tag{3.8}$$

where $S = \sum_n |e[n]|^2$.

This is known as Prony's method for time domain IIR filter design. This method was used to model the test data as the impulse response of an IIR filter. This impulse response is then extrapolated over a longer duration and then its periodogram is computed.

B. NONCAUSAL EXTRAPOLATION

Figure 29 shows the original Kay and Marple data sequence and the 64 point Prony's impulse response estimate of this data sequence. The order of the numerator and denominator in the transfer function of Eq. (3.2) was chosen to be 16. Several model orders were tried before settling on the order of 16. If too low an order is chosen, there is not enough resolution in the periodogram of the extrapolated sequence. If too high an order is chosen, spurious spectral peaks arise. Examples of varying model orders are given in Appendix B. As can be seen in Figure 29, Prony's impulse response estimate of the data sequence for an order of 16 is a very good approximation.

By choosing the appropriate impulse duration, the impulse response can be extrapolated as far as we like. The impulse response can also be extrapolated in both forward and backward directions. We can extend it backwards by reversing the original data sequence and performing the same computations as one normally would for the forward case. Once this new Prony's estimate has been computed, we reverse the extension and attach it to the front of the original data sequence. This gives a noncausal extrapolation of the original data sequence.

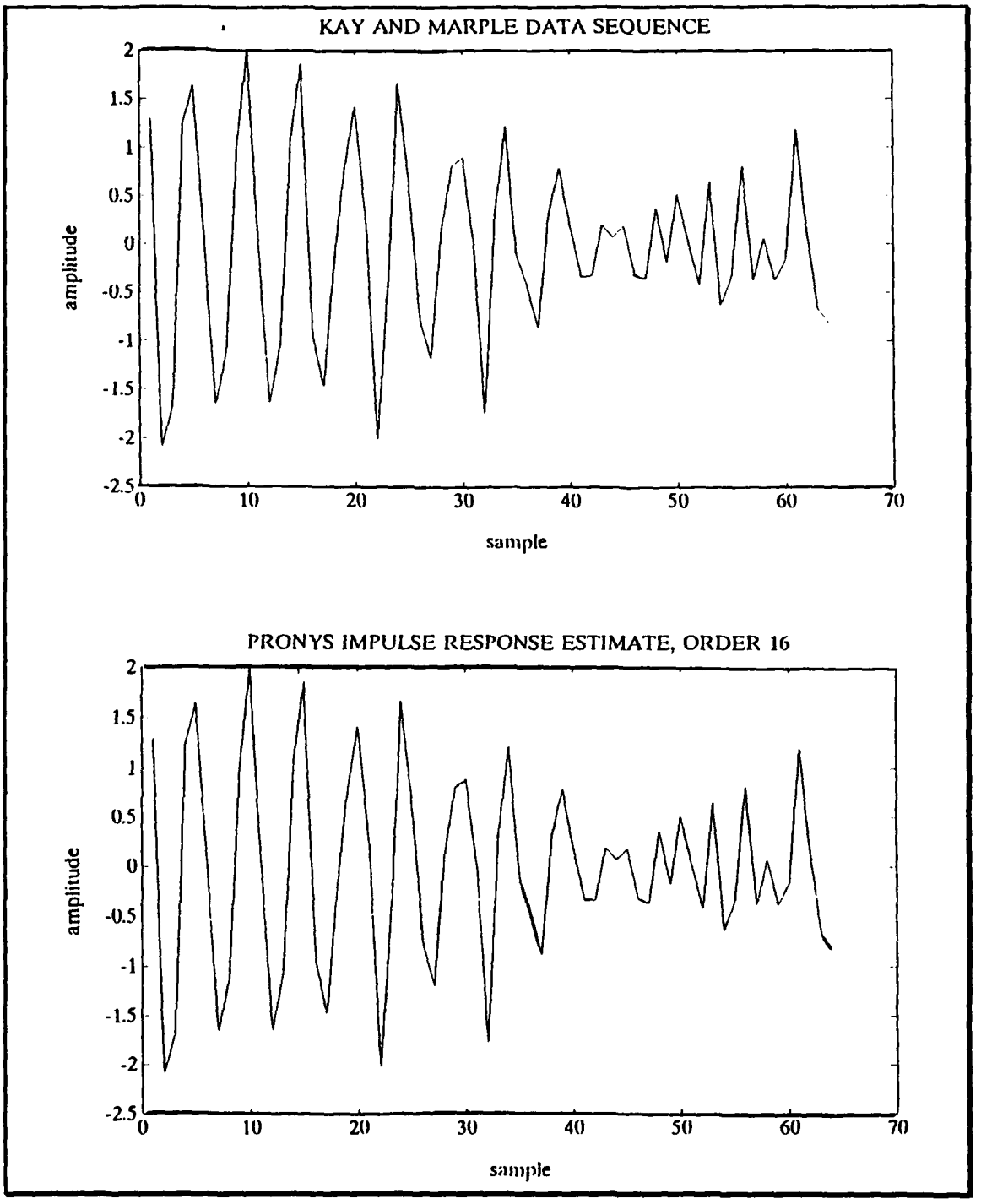

Figure 29. Original KM Data and Prony's Impulse Response Estimate of KM Data

This was the method performed on all test sequences. A strictly causal or forward extrapolation was tried but the results were not as good as the noncausal extrapolation. Some examples of the causal extrapolation are shown in Appendix B. The length of the extrapolation is provided on each figure.

C. RESULTS

For the KM data, the noncausal extrapolation method gave excellent results. Figure 30 shows the extrapolated time series and the resulting periodogram. The three sinusoids present are correctly identified and the colored noise passband process is properly represented.

Figure 31 shows the extrapolated sequence and corresponding periodogram of a test sequence consisting of two sinusoids embedded in white noise. The two sinusoids have frequencies which are 10 Hz apart (f_s = 1000 Hz) and both have zero initial phase. The SNR is 15 dB. The periodogram of the extrapolated sequence correctly identifies the two frequencies present. When the initial phase of one of the sine waves is changed to 45 degrees, the periodogram of the extrapolated sequence once again identifies two frequencies, however one of the frequency estimates is slightly off. Figure 32 shows the results. Figure 33 shows the method does not distinguish the two frequencies when one of the initial phase angles is changed to 90 degrees.

Figures 34 through 36 show the extrapolated time series and resulting periodograms for two sine waves 5 Hz apart in frequency with different initial phase angles. The SNR is 30 dB in all three cases. The periodograms in each of the three cases correctly distinguished the two frequencies present when the initial relative phases were 0 degrees (Figure 34), 45 degrees (Figure 35), and 90 degrees (Figure 36).

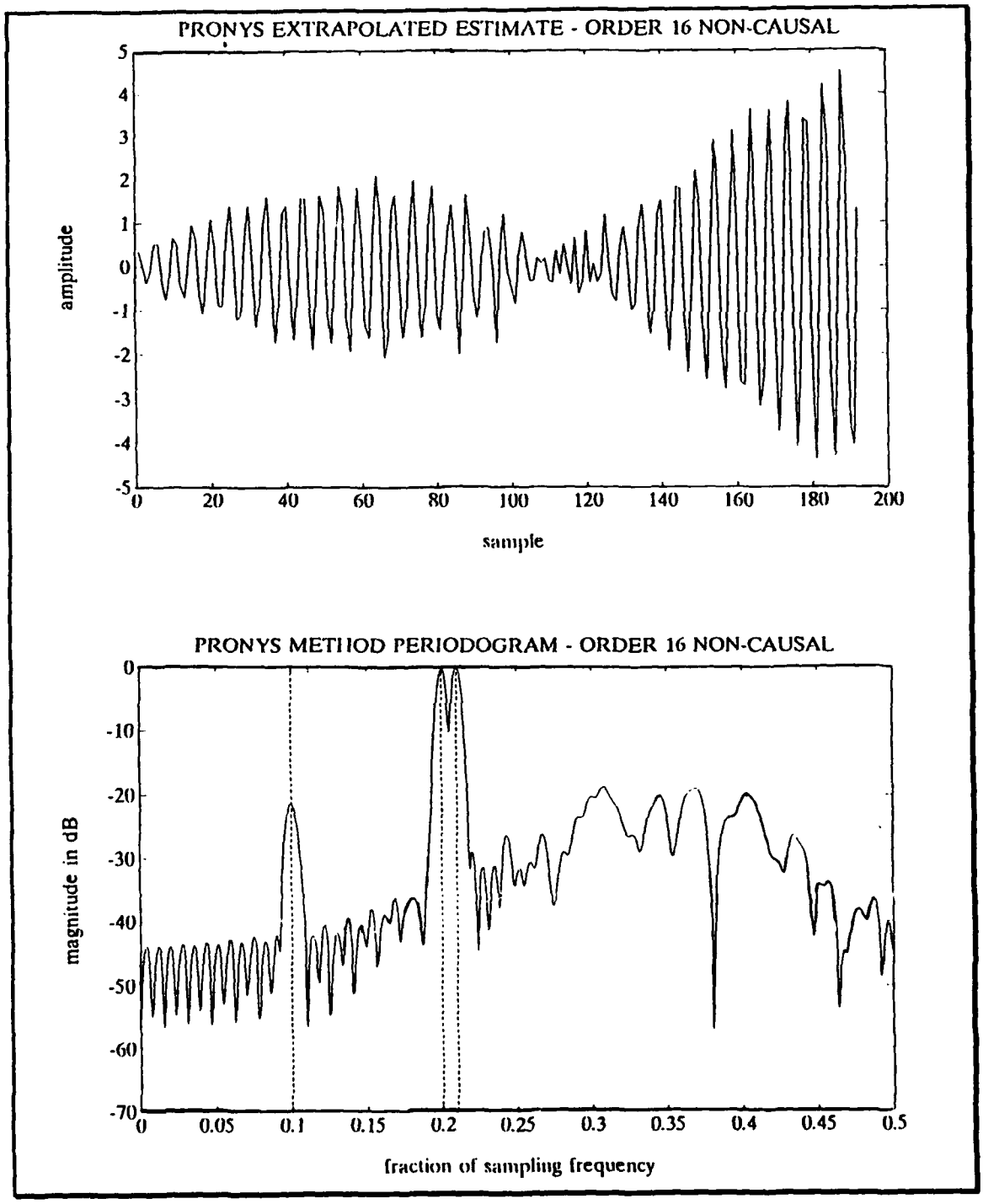

Figure 30. Extrapolated KM Data and Respective Periodogram

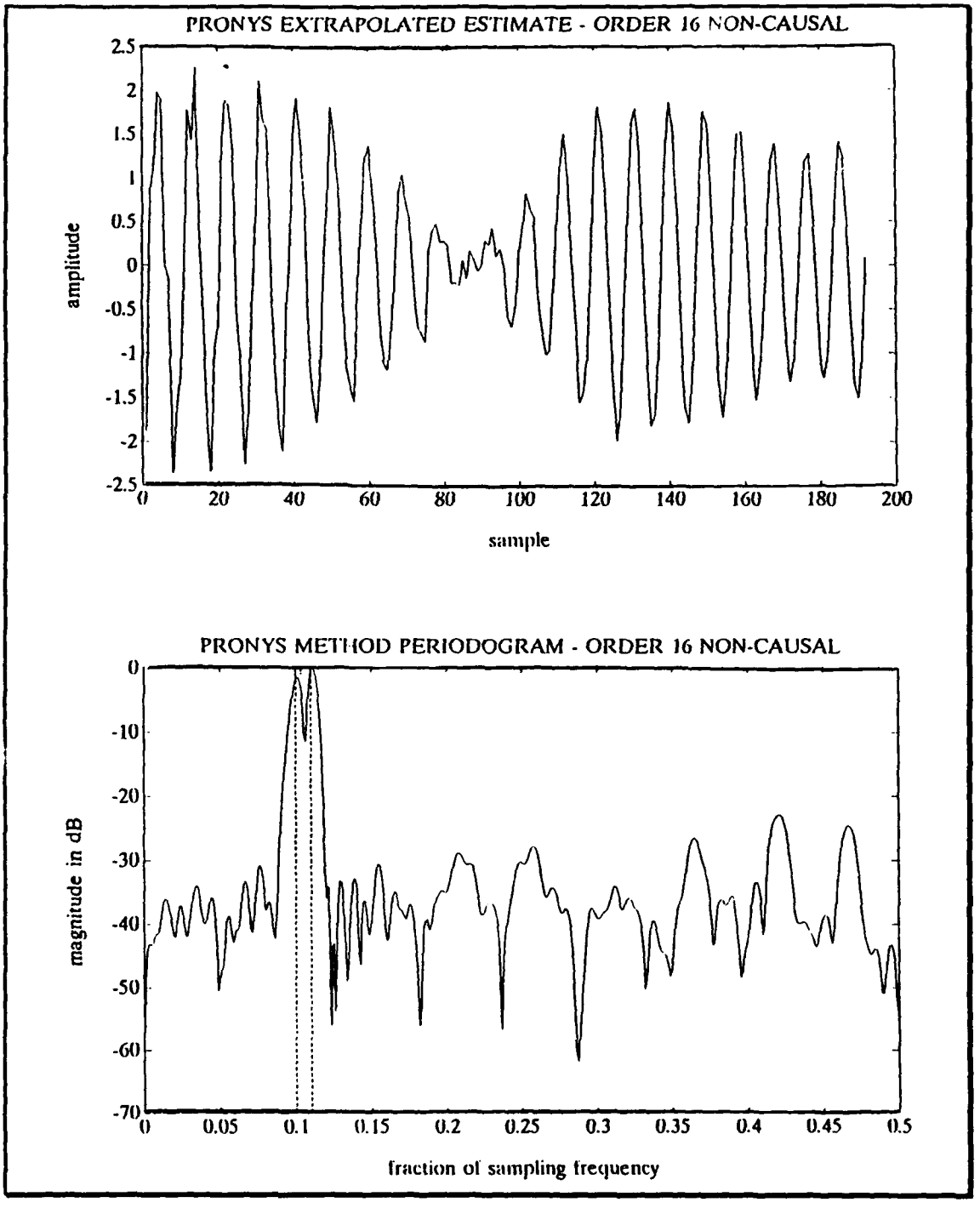

Figure 31. Extrapolated Sequence of Two Sine Waves, 10 Hz Apart, 0° Initial Phase Angle, SNR = 15 dB, and Resulting Periodogram

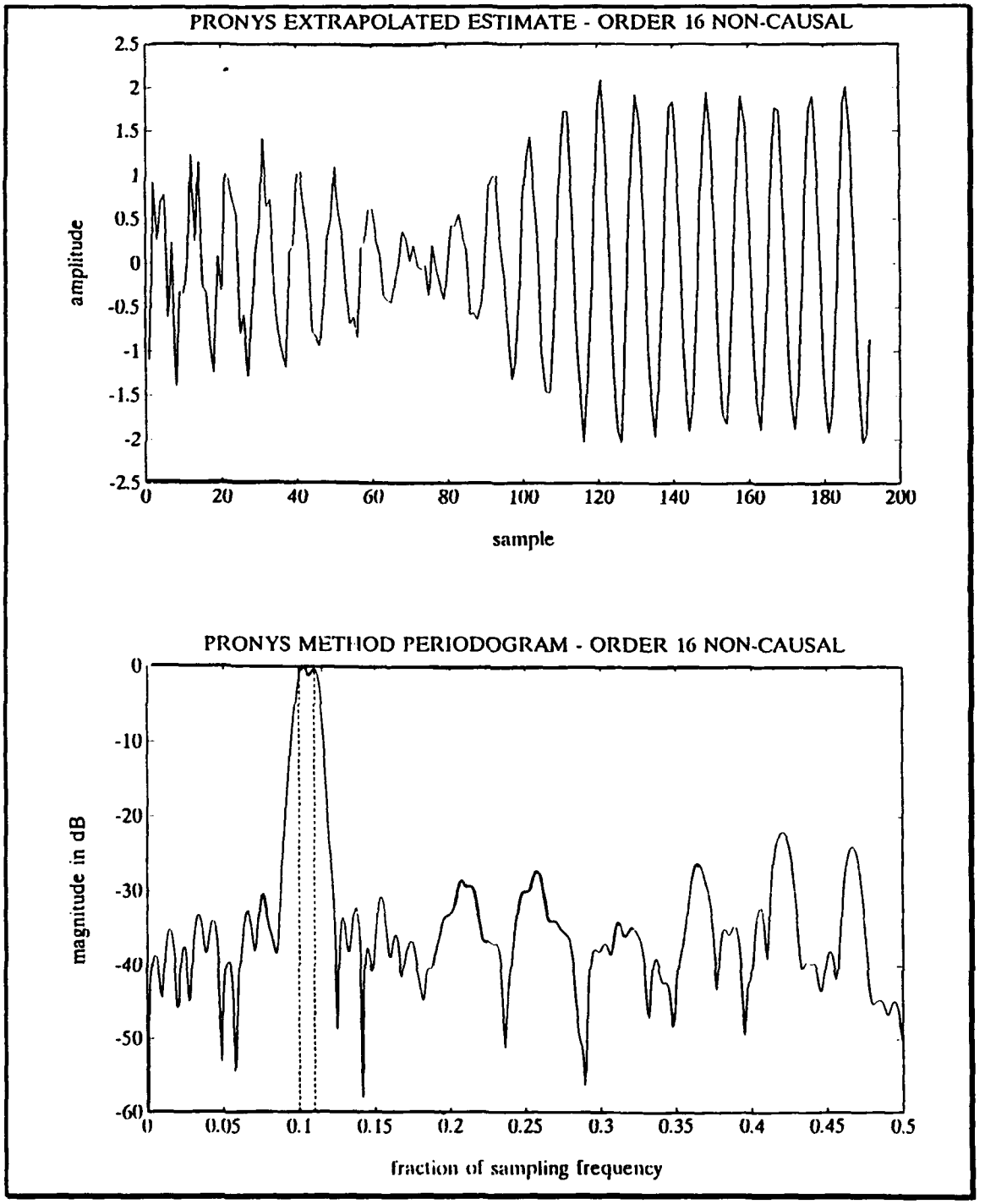

Figure 32. Extrapolated Sequence of two Sine Waves, 10 Hz Apart, 45° Initial Phase Angle, SNR = 15 dB, and Resulting Periodogram

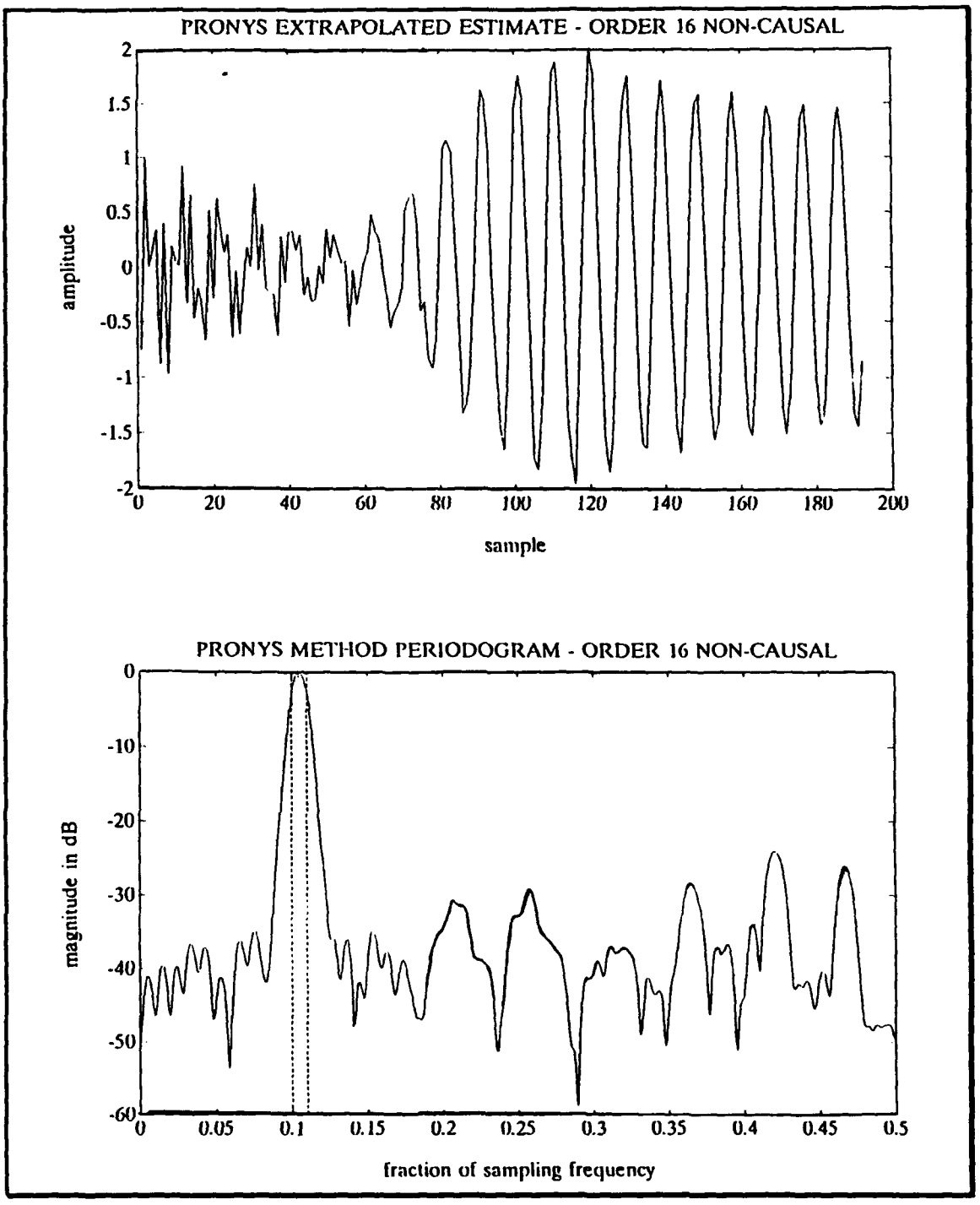

Figure 33. Extrapolated Sequence of two Sine Waves, 10 Hz Apart, 90° Initial Phase Angle, SNR = 15 dB, and Resulting Periodogram

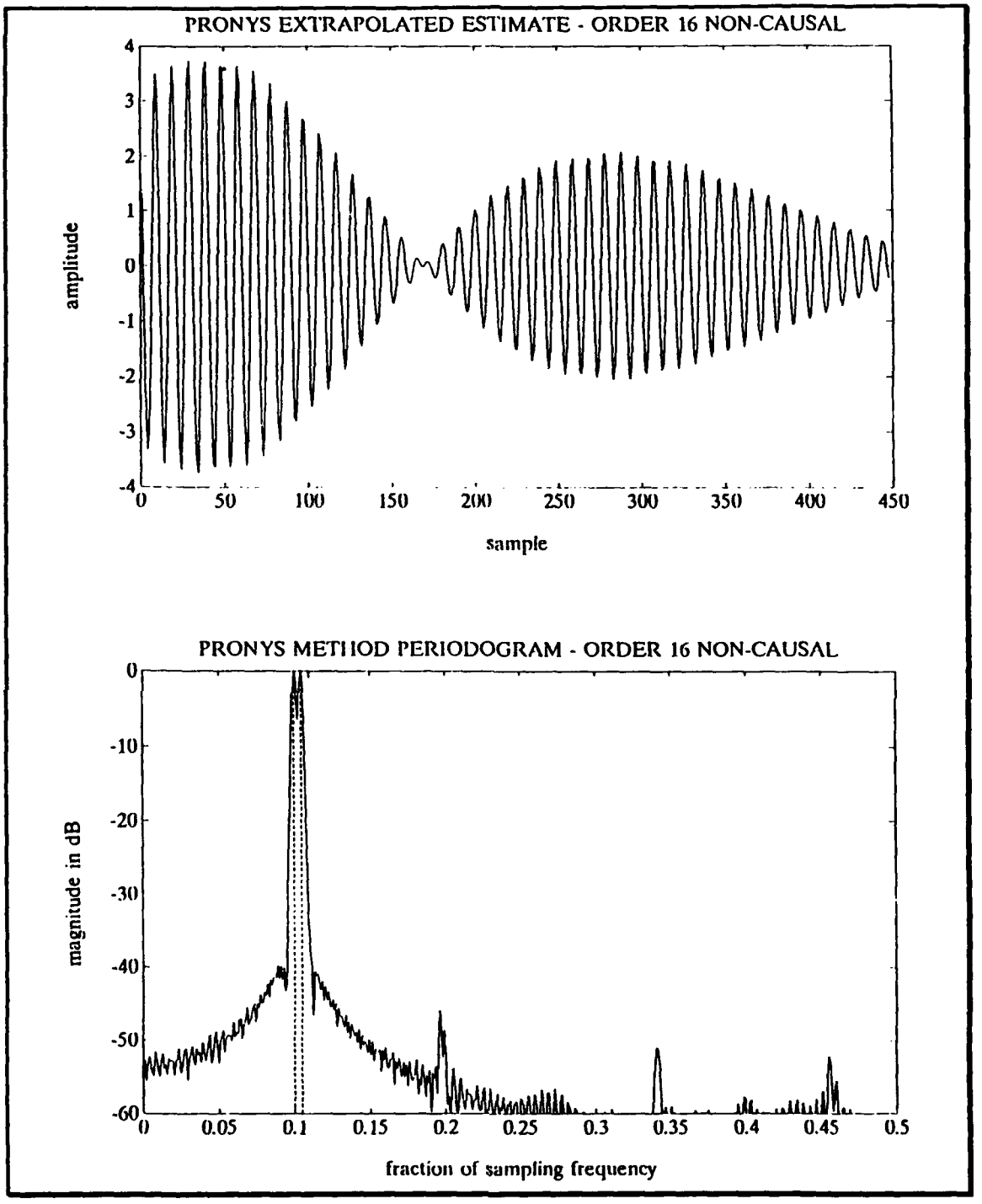

Figure 34. Extrapolated Sequence of two Sine Waves, 5 Hz Apart, 0° Initial Phase Angle, SNR = 30 dB, and Resulting Periodogram

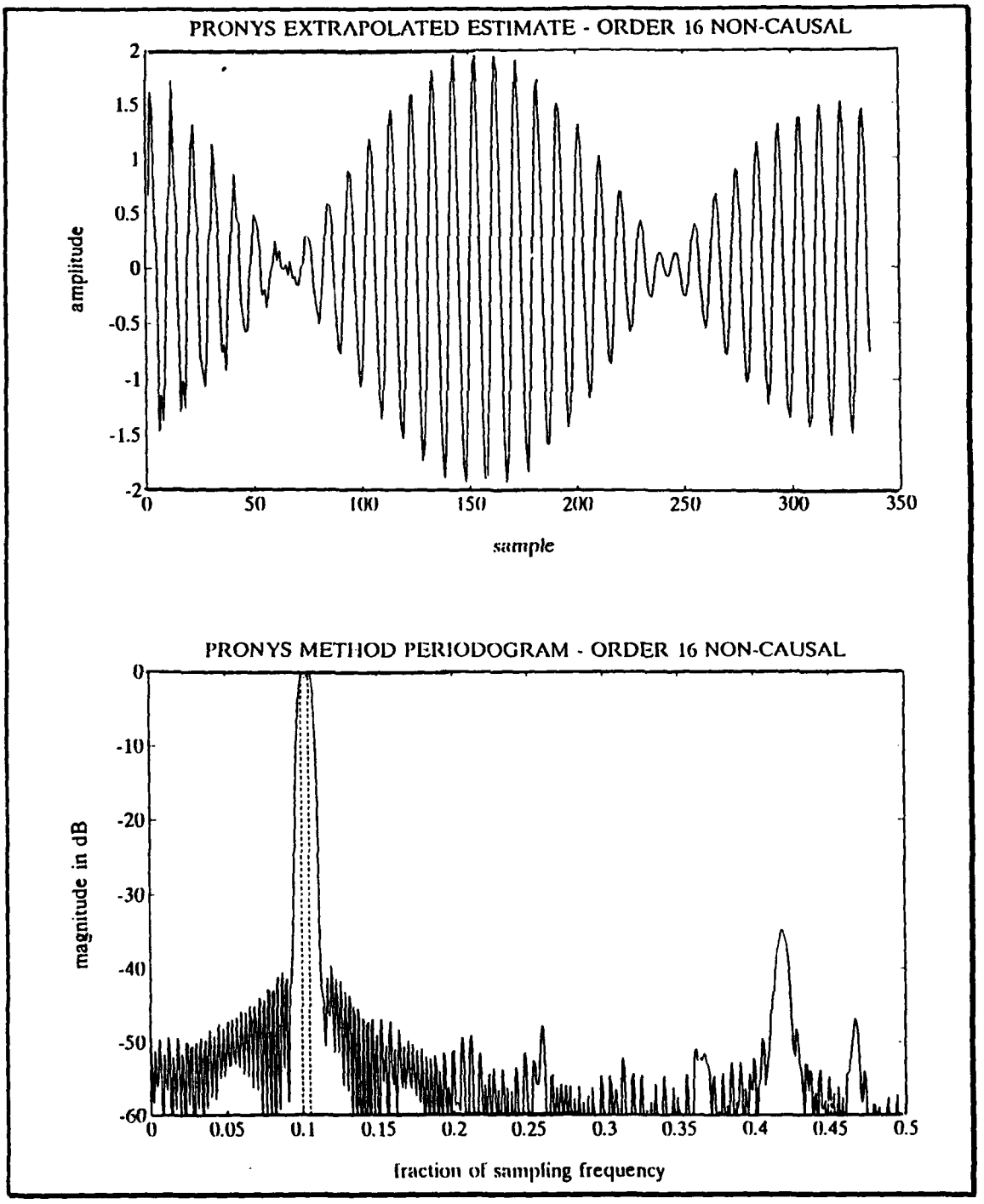

Figure 35. Extrapolated Sequence of two Sine Waves, 5 Hz Apart, 45° Initial Phase Angle, SNR = 30 dB, and Resulting Periodogram

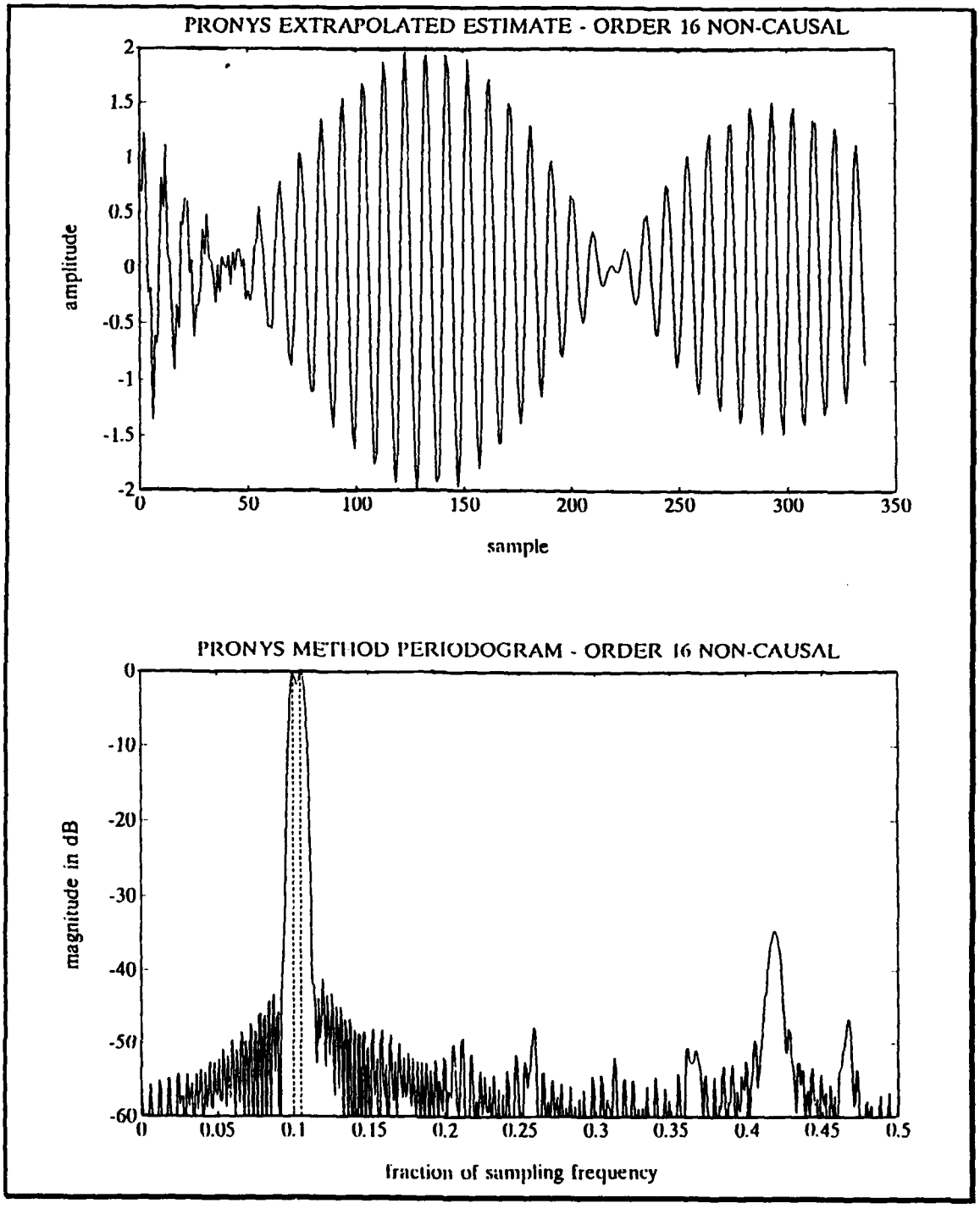

Figure 36. Extrapolated Sequence of two Sine Waves, 5 Hz Apart, 90° Initial Phase Angle, SNR = 30 dB, and Resulting Periodogram

The periodogram resulting from Prony's method of data extrapolation is unable to distinguish the two sinusoids which are 5 Hz apart in a SNR less than approximately 27 dB. Prony's method of data extrapolation resulted in periodograms which were not as accurate as the MFBLP method in lower SNRs. In high SNRs (20 dB or higher) good results are obtained but the periodograms resulting from the data extrapolation using the MFBLP prediction coefficients still provide better estimates.

IV. EIGENVECTOR EXTRAPOLATION

A. DEFINITION

In this chapter we develop a series expansion for a random vector in terms of an arbitrary set of orthonormal vectors [Ref. 8:pp. 261-262]. Once this set of orthonormal vectors (eigenvectors) are computed, we will use the dominant ones to extrapolate the data sequence and, hopefully, improve the spectral resolution.

First consider a set of N orthonormal vectors $\mathbf{q}_1, \mathbf{q}_2, \ldots, \mathbf{q}_N$ that satisfy the following two requirements:

$$\mathbf{q}_k^T \mathbf{q}_i = \begin{cases} 1, & k = i \\ 0, & k \neq i \end{cases} \tag{4.1}$$

Let $\mathbf{x}$ denote an N-by-1 random vector with zero mean and correlation matrix $\mathbf{R}_x$. The requirement is to express the random vector $\mathbf{x}$ as a linear combination of the set of orthonormal vectors $\mathbf{q}_1, \mathbf{q}_2, \ldots, \mathbf{q}_N$ as follows:

$$\mathbf{x} = \sum_{i=1}^{N} \alpha_i \mathbf{q}_i. \tag{4.2}$$

Equations (4.1) and (4.2) define the discrete time version of the Karhunen-Loéve expansion [Ref. 8:p. 261]. The coefficients of the expansion, α_i, are defined by the inner product between the vectors $\mathbf{x}$ and $\mathbf{q}_i$, given by

$$\begin{aligned} \alpha_i &= \mathbf{q}_i^T \mathbf{x} \\ &= \mathbf{x}^T \mathbf{q}_i \quad \text{for } i = 1,2,\ldots,N. \end{aligned} \tag{4.3}$$

The coefficients α_i are obviously random variables and we would like to choose the $\mathbf{q}_i$ so that these coefficients are mutually uncorrelated.

Because the random vector x has zero mean, we observe that all the coefficients α_i also have a mean of zero, that is

$$E[\alpha_i] = 0 \quad \text{for } i = 1,2,\ldots,N. \tag{4.4}$$

Therefore, for the coefficients of expansion to be uncorrelated, we require that they satisfy the condition

$$\begin{aligned} \operatorname{cov}[\alpha_k \alpha_i] &= E\left[(\alpha_k - E[\alpha_k])(\alpha_i - E[\alpha_i])\right] \\ &= E[\alpha_k \alpha_i] \\ &= \begin{cases} \upsilon_i, & k = i \\ 0, & k \neq i \end{cases} \end{aligned} \tag{4.5}$$

where the υ_i are constants to be specified. From Eq. 4.3 we can rewrite Eq. 4.5 as

$$\begin{aligned} \operatorname{cov}[\alpha_k \alpha_i] &= E[\alpha_k \alpha_i] \\ &= E\left[\mathbf{q}_k^T \mathbf{x}\mathbf{x}^T \mathbf{q}_i\right] \\ &= \mathbf{q}_k^T E\left[\mathbf{x}\mathbf{x}^T\right] \mathbf{q}_i \end{aligned} \tag{4.6}$$

Since by definition the correlation matrix of the random vector x equals

$$\mathbf{R}_x = E\left[\mathbf{x}\mathbf{x}^T\right], \tag{4.7}$$

we can simplify Eq. 4.6 as

$$\operatorname{cov}[\alpha_k \alpha_i] = \mathbf{q}_k^T \mathbf{R}_x \mathbf{q}_i. \tag{4.8}$$

Therefore from Eq. 4.5 and Eq. 4.8 we have

$$\mathbf{q}_k^T \mathbf{R}_x \mathbf{q}_i = \begin{cases} \upsilon_i, & k = i, \\ 0 & k \neq i \end{cases} \tag{4.9}$$

If we multiply both sides of Eq. 4.1 by υ_i, we have

$$\upsilon_i \mathbf{q}_k^T \mathbf{q}_i = \begin{cases} \upsilon_i, & k = i, \\ 0 & k \neq i \end{cases} \tag{4.10}$$

Therefore for Eq. 4.9 to hold for all choices of k and a particular i, it is necessary that the matrix product $\mathbf{R}_x\mathbf{q}_i$ equal $\upsilon_i\mathbf{q}_i$, as shown by

$$\mathbf{R}_x\mathbf{q}_i = \upsilon_i\mathbf{q}_i \tag{4.11}$$

where the scalar υ_i is the eigenvalue of the correlation matrix $\mathbf{R}_x$ and the vector $\mathbf{q}_i$ is the eigenvector associated with the eigenvalue υ_i.

Therefore we can express the random vector x as a linear combination of the eigenvectors of the correlation matrix of x. Once we have computed the eigenvalues and eigenvectors, we will extract the dominant ones and use them to extend the original data sequence by appending this new sequence to the original sequence. All test sequences are extrapolated to a length of 128 points.

B. RESULTS

The first set of test data consists of two sinusoids with a 10 Hz (f_s = 1000 Hz) difference in frequency and 0 degree relative phase angle. By examining the eigenvalues of the correlation matrix, we observe that there are two dominant eigenvectors. This corresponds to the fact that we have two sinusoids present. Because there are two dominant eigenvectors present, we use the linear combination of the two most dominant eigenvectors for the extrapolation. Therefore in Eq. 4.2, N is set to 2 and the resulting sequence is appended to the original data. The autocorrelation matrix in Eq. 4.2 is computed by using the biased autocorrelation method. Figure 37 shows the plot of the extrapolated data sequence and the resulting periodogram. Two distinct peaks are present but they are slightly off. The SNR is 15 dB.

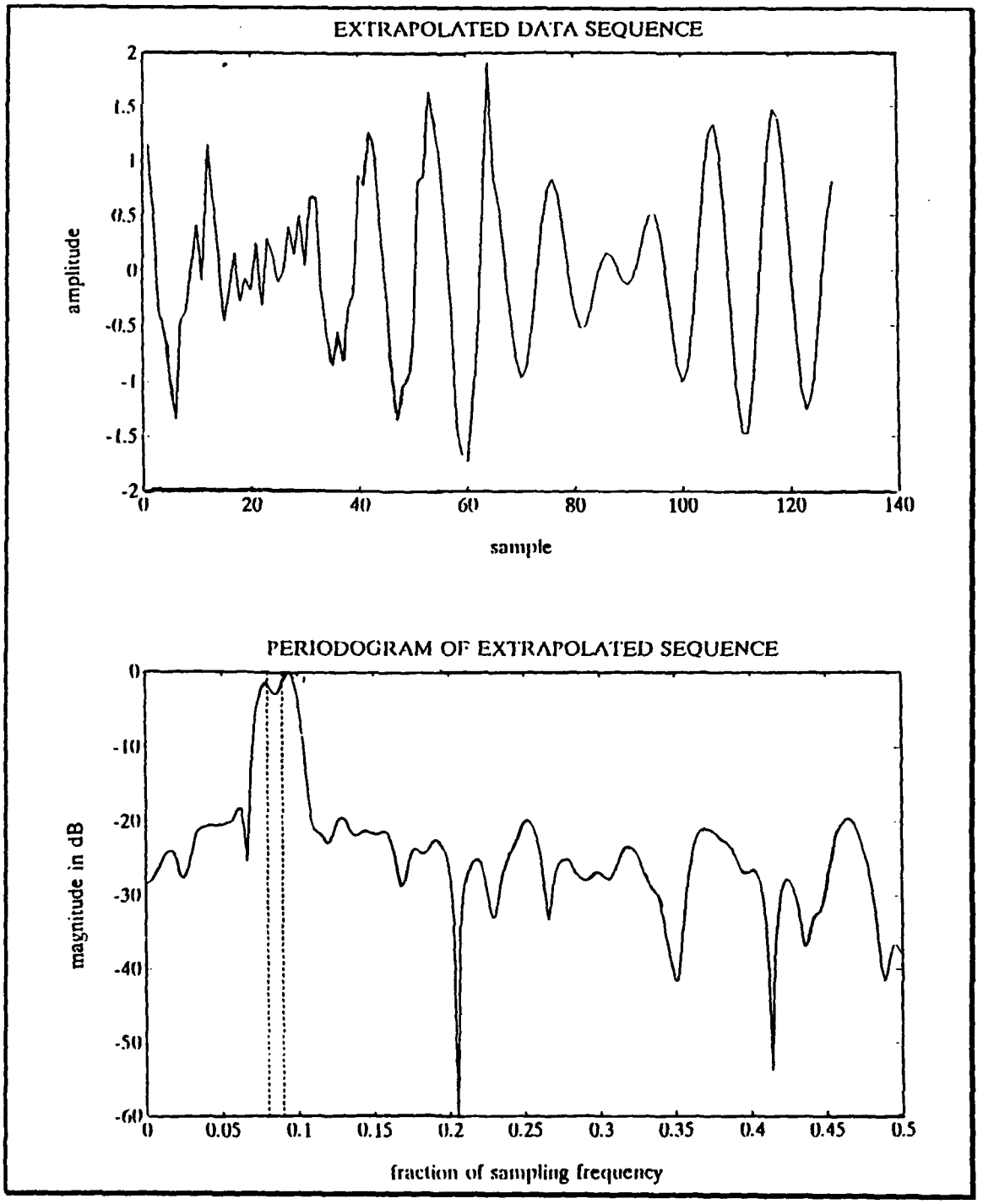

Figure 37. Eigenvector Extrapolation of 2 Sine Waves, 10 Hz Apart, 0° Initial Phase Angle, SNR = 15 dB, 2 Eigenvectors Used; and Resulting Periodogram

Figure 38 shows the results when the initial relative phase angle between the two sine waves is changed to 45 degrees. Once again two distinct spectral peaks are present but they are slightly off from the true values. Figure 39 shows the results when the relative phase angle is changed to 90 degrees. As in the two prior cases, the spectral peaks are slightly off from their true values.

Even though the spectral peaks are not absolutely accurate, they remain fairly consistent in even lower SNRs. Figure 40 shows the results of the two sine waves (initially in phase, 10 Hz apart) in a SNR of -3 dB. The spectral peaks are still well defined. When the SNR was lowered to -6 dB, the periodogram was unable to distinguish the two peaks. Figure 41 shows these results.

Figure 42 shows the results of the eigenvector extrapolation on the KM sequence. N was chosen to be 3 as we have three sinusoids present in this case. Three main spectral peaks are present but one of the estimates is slightly off.

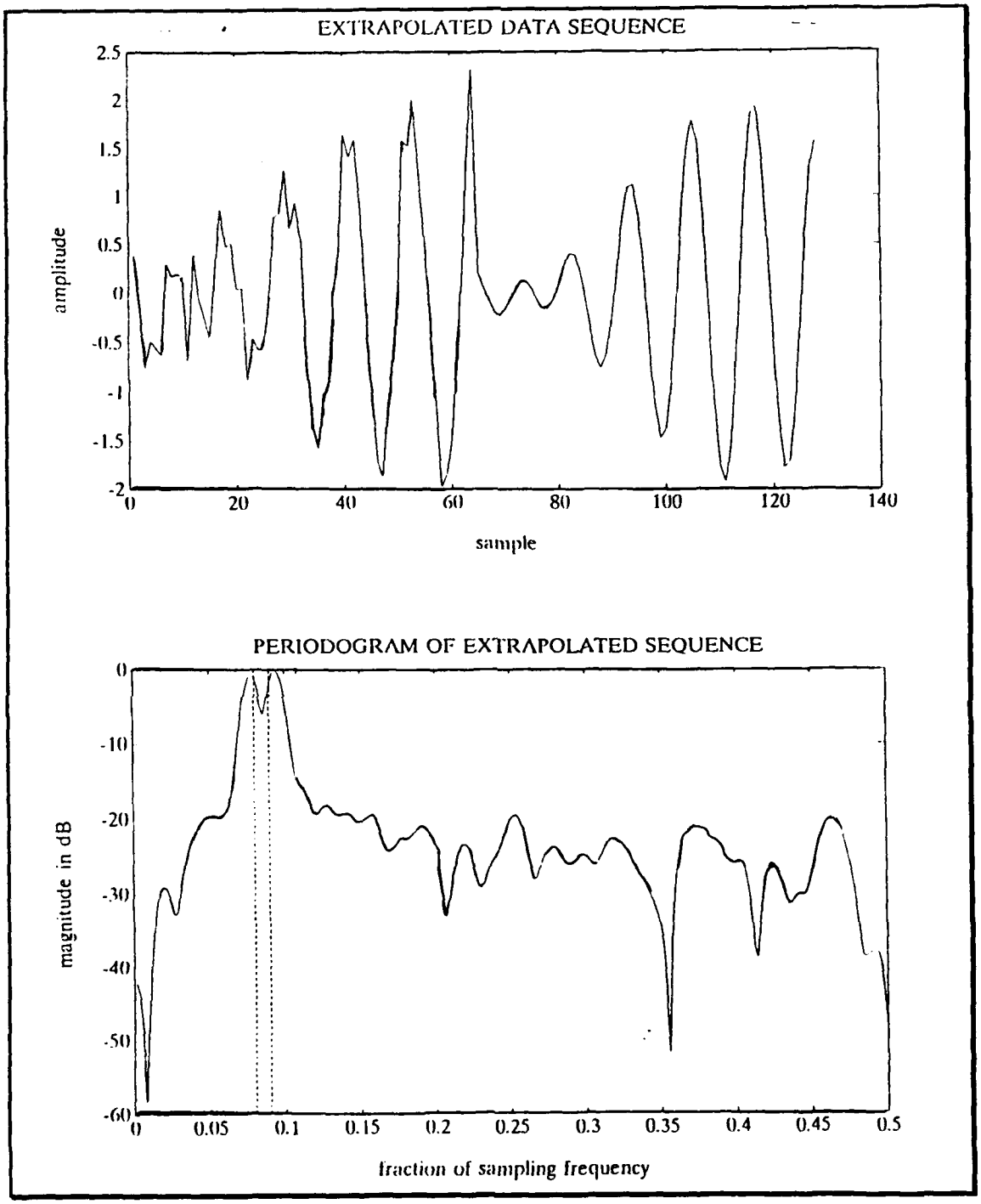

Figure 38. Eigenvector Extrapolation of Two Sine Waves, 10Hz Apart, 45° Initial Phase Angle, SNR = 15 dB, 2 Eigenvectors Used; and Resulting Periodogram

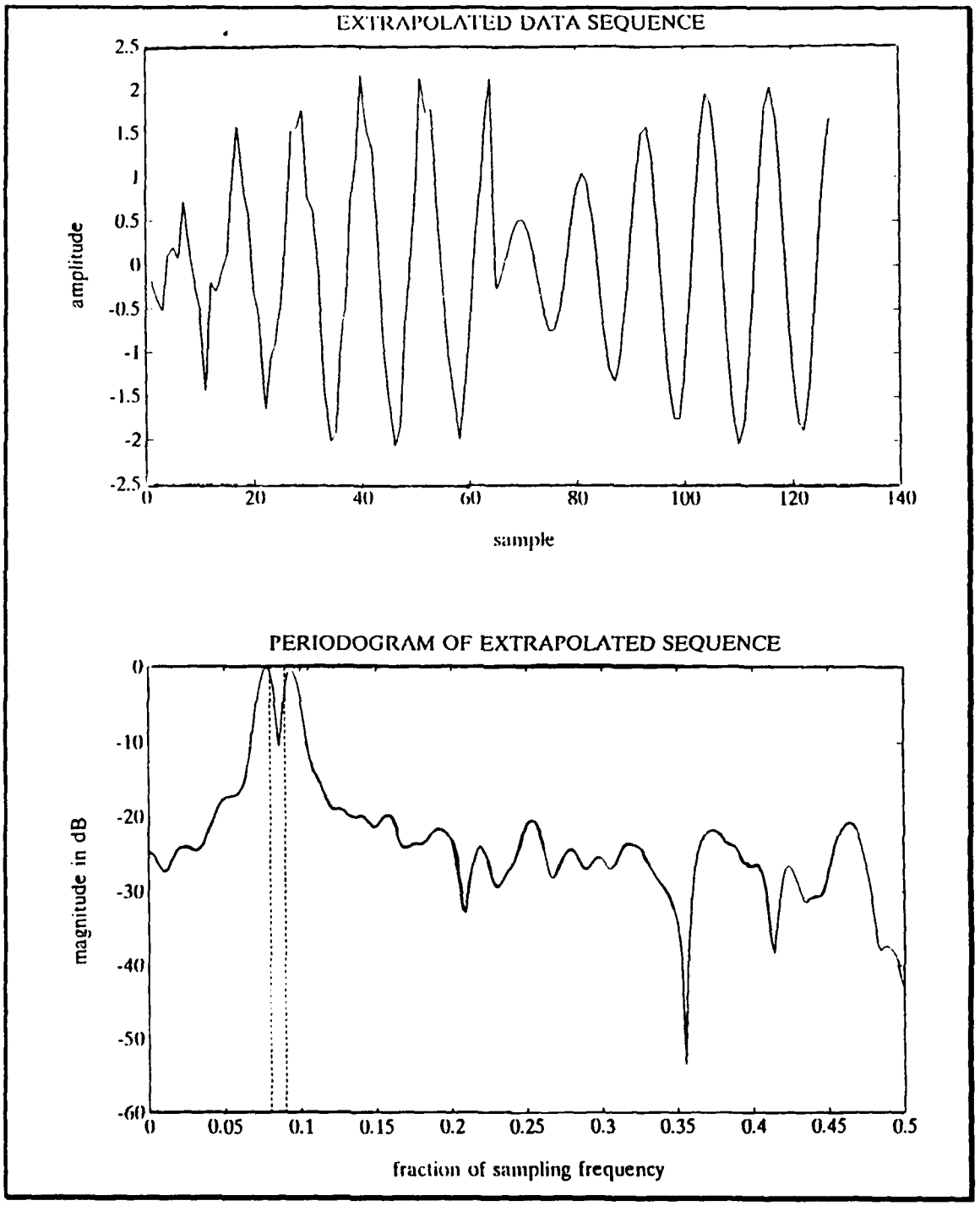

Figure 39. Eigenvector Extrapolation of 2 Sine Waves, 10 Hz Apart, 90° Initial Phase Angle, SNR = 15 dB, 2 Eigenvectors used; and Resulting Periodogram

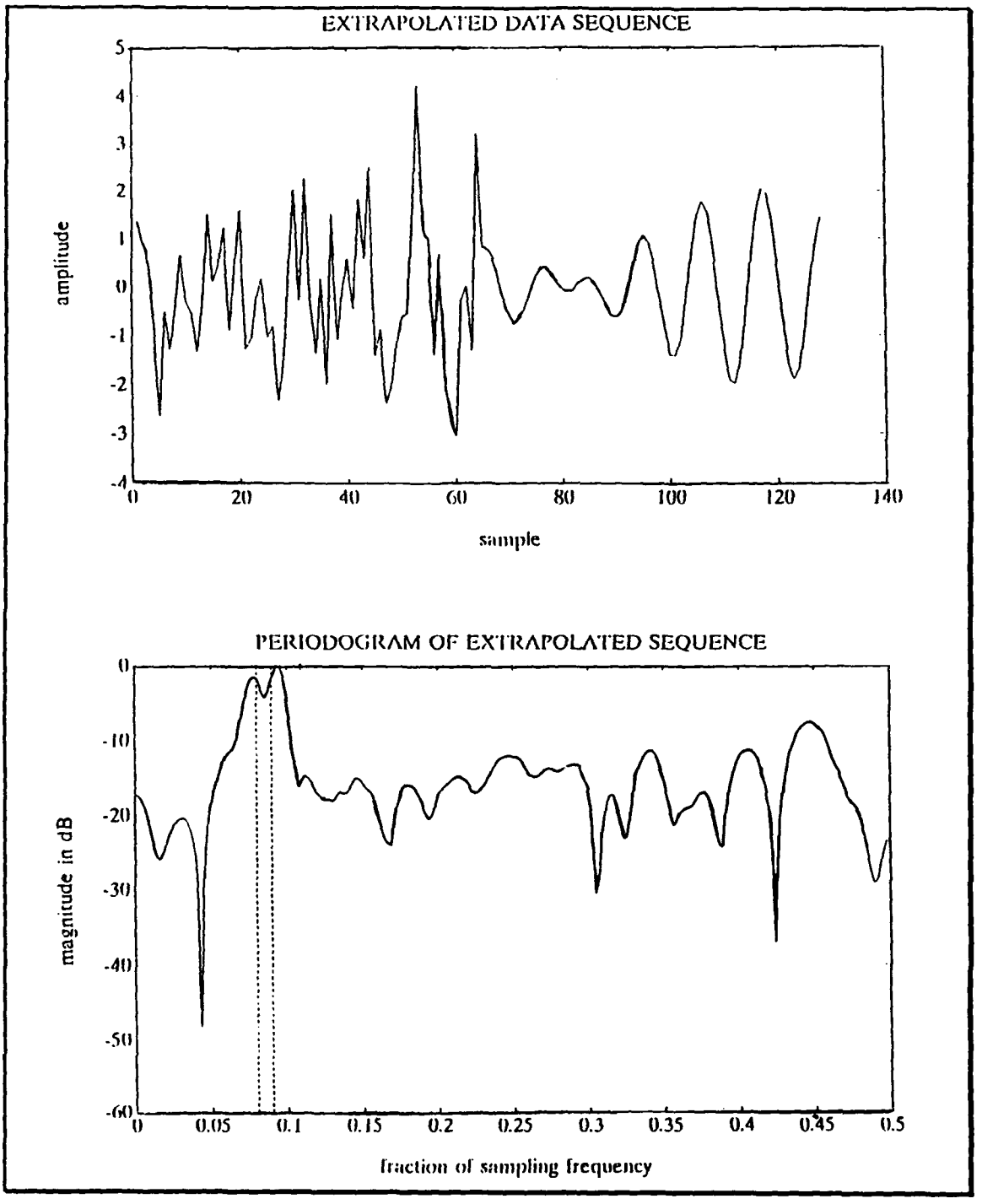

Figure 40. Eigenvector Extrapolation of 2 Sine Waves, 10 Hz Apart, 0° Initial Phase Angle, SNR = –3 dB, and Resulting Periodogram

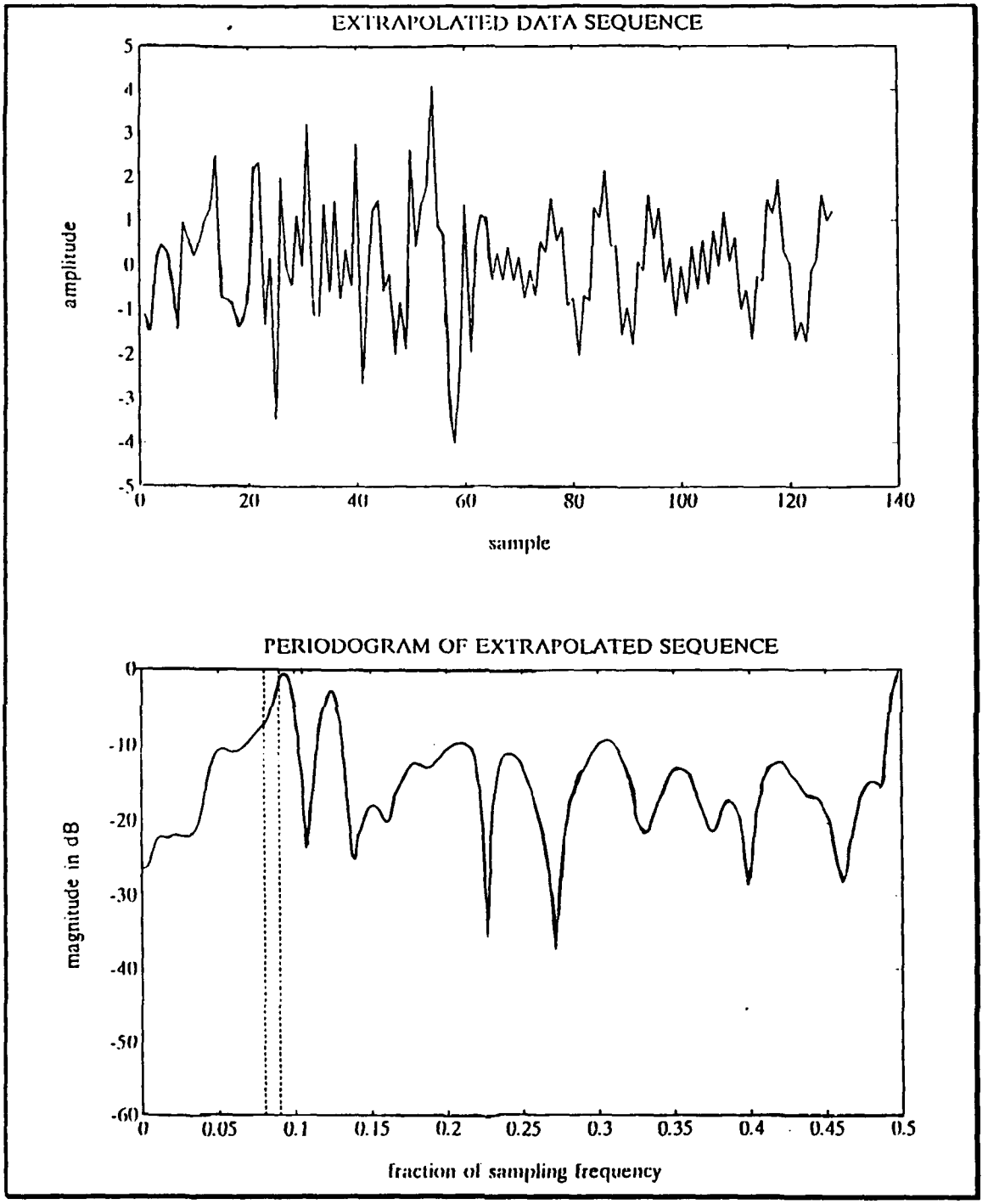

Figure 41. Eigenvector Extrapolation of 2 Sine Waves, 10 Hz Apart, 0° Initial Phase Angle, SNR = –6 dB, and Resulting Periodogram

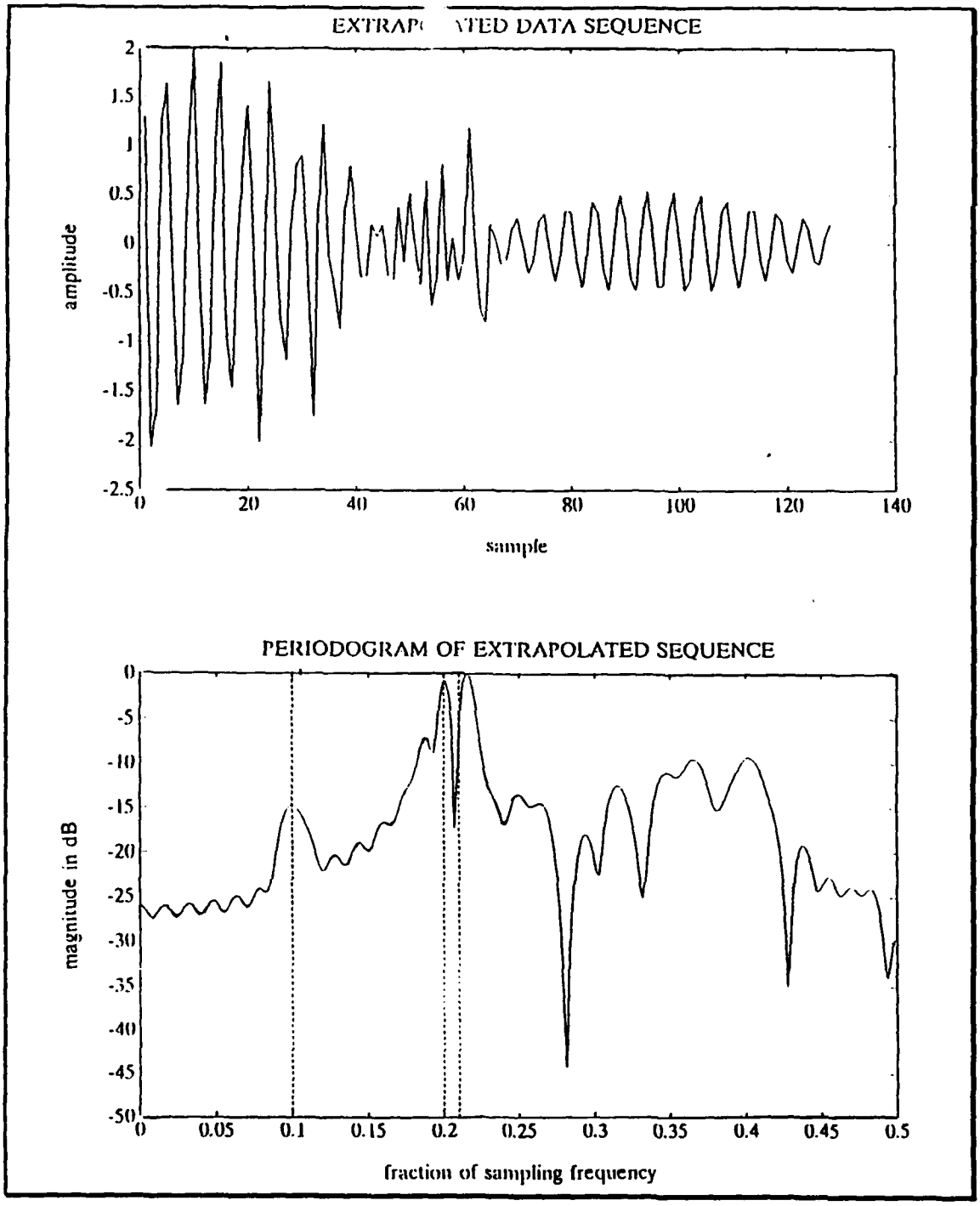

Figure 42. Eigenvector Extrapolation of KM Data, 3 Eigenvectors Used, and Resulting Periodogram

V. CONCLUSIONS AND RECOMMENDATIONS

In this thesis, data extrapolation techniques to enhance the spectral resolution of the periodogram are presented. The three methods of extrapolation are: linear predictive filtering, modeling the data as the impulse response of a filter (Prony's method), and eigenvector extrapolation using the Karhunen-Loéve expansion.

The most promising of the three methods is the linear predictive filtering extrapolation using the prediction coefficients obtained from the modified forward-backward method. This technique resulted in spectral estimates which were equivalent to or better than those obtained via the autoregressive method. Frequencies as closely spaced as 2 Hz have been resolved, with an appropriate SNR, where normally resolution of no better than 12.5 Hz would be obtained using a periodogram of the nonextrapolated sequence. This method performed very well in low SNRs.

In a high SNR environment, the periodograms resulting from Prony's method of extrapolation gave good results. Performance was not as promising in lower SNRs. Model order selection in Prony's method proved to be very important in determining the resolution of the resulting periodogram.

None of the methods covered required very long computer run times. A typical run lasted approximately 1 to 2 minutes on a 286 based personal computer, depending on the extrapolation length. As discussed earlier, care must be taken when extrapolating the data sequence to ensure that it does not grow without bound. This has devastating effects on the spectrum estimation.

In this thesis we have shown that data extrapolation is a viable means for increasing the resolution of the periodogram. Further research should concentrate on other data extrapolation techniques for better and more robust spectral resolution.

APPENDIX A. OTHER METHODS OF AR PARAMETER ESTIMATION

A. AUTOCORRELATION METHOD

In the autocorrelation method or the Yule-Walker approach, the AR parameters are estimated by minimizing an estimate of the prediction error power

$$\hat{\rho} = \frac{1}{N} \sum_{n=-\infty}^{\infty} \left| x[n] + \sum_{k=1}^{P} a[k]x[n-k] \right|^2 . \tag{A.1}$$

The samples of x[n] which are not observed, those not in the range $0 \le n \le N-1$, are set equal to zero. To minimize the prediction error power, we differentiate Eq. A.1 with respect to the real and imaginary parts of the a[k]s. This yields

$$\frac{1}{N} \sum_{n=-\infty}^{\infty} \left(x[n] + \sum_{k=1}^{P} a[k]x[n-k] \right) x^*[n-\ell] = 0, \quad \ell = 1,2,\ldots P. \tag{A.2}$$

In matrix form this set of equations becomes

$$\begin{bmatrix} r_{xx}[0] & r_{xx}[-1] & \cdots & r_{xx}[-(P-1)] \\ r_{xx}[1] & r_{xx}[0] & \cdots & r_{xx}[-(P-2)] \\ \vdots & & & \\ r_{xx}[P-1] & r_{xx}[P-2] & \cdots & r_{xx}[0] \end{bmatrix} \begin{bmatrix} a[1] \\ a[2] \\ \vdots \\ a[3] \end{bmatrix} = - \begin{bmatrix} r_{xx}[1] \\ r_{xx}[2] \\ \vdots \\ r_{xx}[P] \end{bmatrix} \tag{A.3}$$

where

$$r_{xx}[k] = \begin{cases} \frac{1}{N} \sum_{n=0}^{N-1-k} x^*[n]x[n+k] & \text{for } k = 0,1,\ldots,P \\ r_{xx}^*[-k] & \text{for } k = -(p-1),-(p-2),\ldots,-1 \end{cases} \tag{A.4}$$

which is recognized as the biased autocorrelation function (ACF) estimator. The AR parameters or prediction coefficients are easily solved for from Eq. A.3.

The matrix in Eq. A.3 is hermitian ($r_{xx}[-k] = r^{*}_{xx}[k]$) and Toeplitz, and furthermore can be shown to be positive definite. As such, the resulting estimated poles are guaranteed to be within the unit circle. This property ensures that the prediction coefficients calculated from Eq. A.3 will give stable extrapolations. This is very important because if the extrapolation grows without bound, it will have devastating effects on the spectrum estimation.

The autocorrelation method extrapolation was found to produce poorer resolution of the spectral estimates than the other extrapolation techniques used. The length of the extrapolated sequences in this section is 320 points. Figure 43 shows the resulting periodogram of the autocorrelation method extrapolation and the respective AR spectrum estimation of the KM data. The extrapolation gave a periodogram which correctly identifies the 3 sinewaves present but the spectral peaks are not as well defined as in the periodogram obtained from the MFBLP extrapolation technique. The AR spectrum estimation resulting from the autocorrelation method is unable to distinguish the two closely spaced sinusoids at 0.2 and 0.21.

When the autocorrelation method extrapolation was applied to a data sequence consisting of two sine waves embedded in white noise (SNR=15 dB), the resulting periodogram was unable to distinguish the two closely spaced sinusoids (5 Hz apart, f_s = 1000 Hz). The AR spectrum estimation also failed. Figure 44 shows the results.

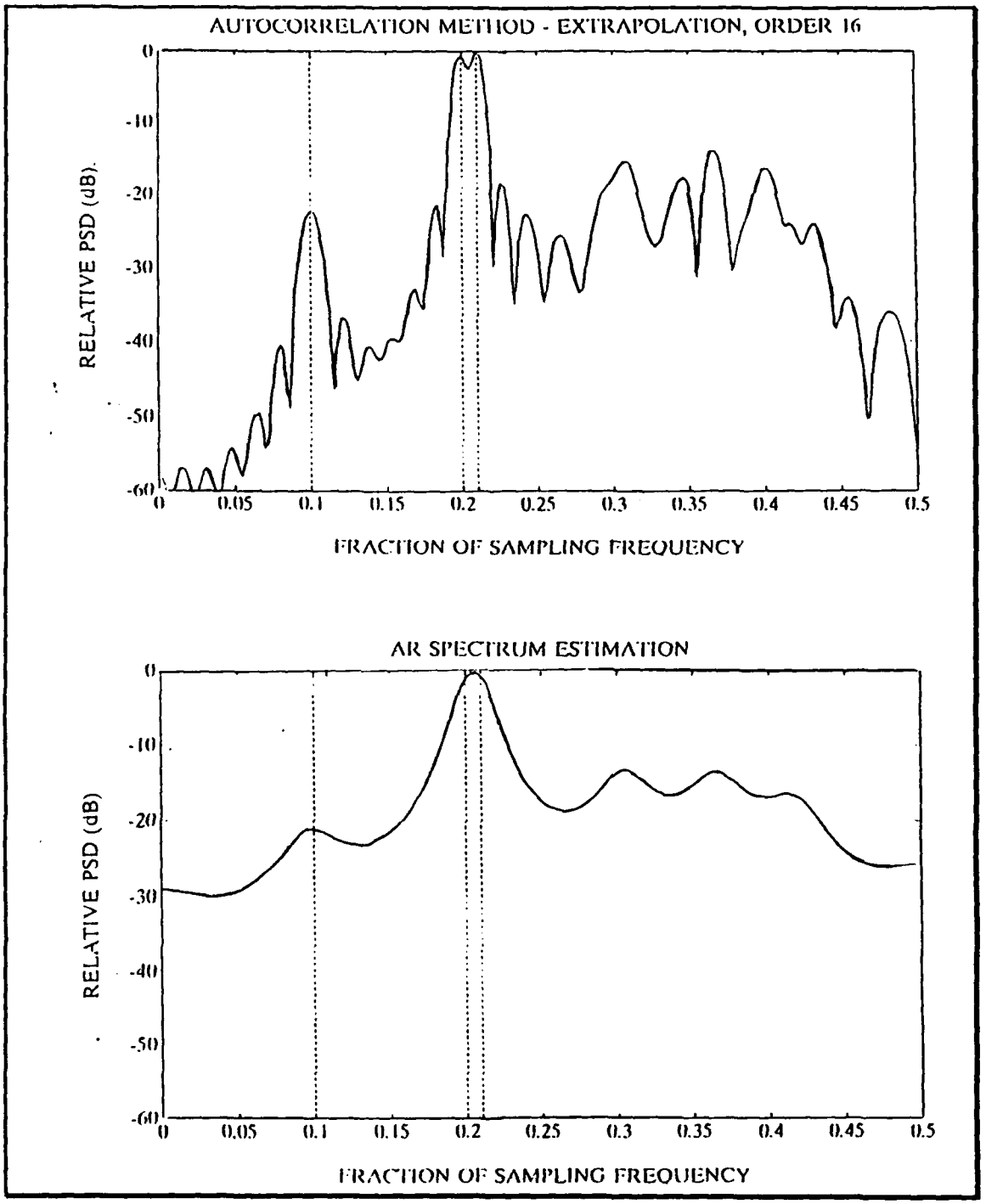

Figure 43. Periodogram of Extrapolated Sequences and AR Spectral Estimation of KM Data Using the Autocorrelation Method

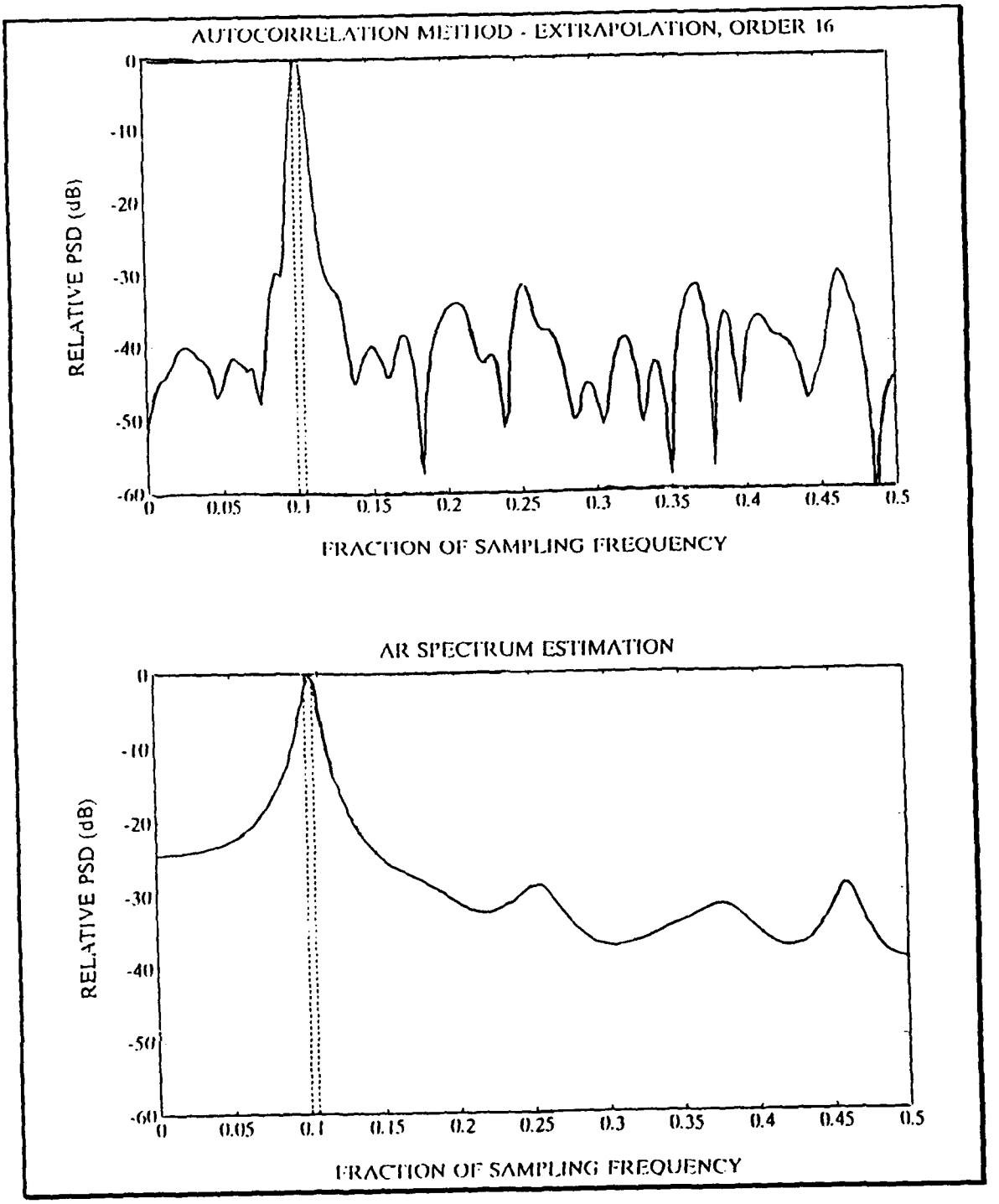

Figure 44. Periodogram Resulting from Extrapolation of 2 Sine Waves, 5 Hz Apart, SNR = 15 dB, 0° Initial Phase Angle and Respective AR Spectrum Estimation

B. COVARIANCE METHOD

In the covariance method, the AR parameters are estimated by minimizing the estimate of the prediction error power

$$\hat{\rho} = \frac{1}{N-P}\sum_{n=P}^{N-1}\left|x[n] + \sum_{k=1}^{P} a[k]x[n-k]\right|^2 . \tag{A.5}$$

Note that the only difference between the covariance method and the autocorrelation method is the range of summation in the prediction error power estimate. The minimization of Eq. A.5 yields

$$\begin{bmatrix} c_{xx}[1,1] & c_{xx}[1,2] & \cdots & c_{xx}[1,P] \\ c_{xx}[2,1] & c_{xx}[2,2] & \cdots & c_{xx}[2,P] \\ \vdots & \vdots & & \vdots \\ c_{xx}[P,1] & c_{xx}[P,2] & \cdots & c_{xx}[P,P] \end{bmatrix} \begin{bmatrix} a[1] \\ a[2] \\ \vdots \\ a[P] \end{bmatrix} = -\begin{bmatrix} c_{xx}[1,0] \\ c_{xx}[2,0] \\ \vdots \\ c_{xx}[P,0] \end{bmatrix} \tag{A.6}$$

where

$$c_{xx}[j,k] = \frac{1}{N-P}\sum_{n=P}^{N-1} x^*[n-j]x[n-k]. \tag{A.7}$$

The correlation matrix in Eq. A.6 is positive semidefinite and not Toeplitz. As such, the estimated poles using the covariance method are not guaranteed to lie within the unit circle. Therefore the prediction coefficients are not guaranteed to give bounded extrapolations although in practice the extrapolations are generally bounded.

Results from the covariance method extrapolation were better than the autocorrelation method extrapolation. Once again the data is extrapolated to a length of 320 points. Figure 45 shows the periodogram of the covariance method extrapolation and the respective AR spectrum estimation of the KM data. The periodogram from the covariance method extrapolated data sequence has well defined spectral peaks and correctly identifies the 3 sinusoids presen: The AR spectrum estimation also identifies the 3 sinusoids.

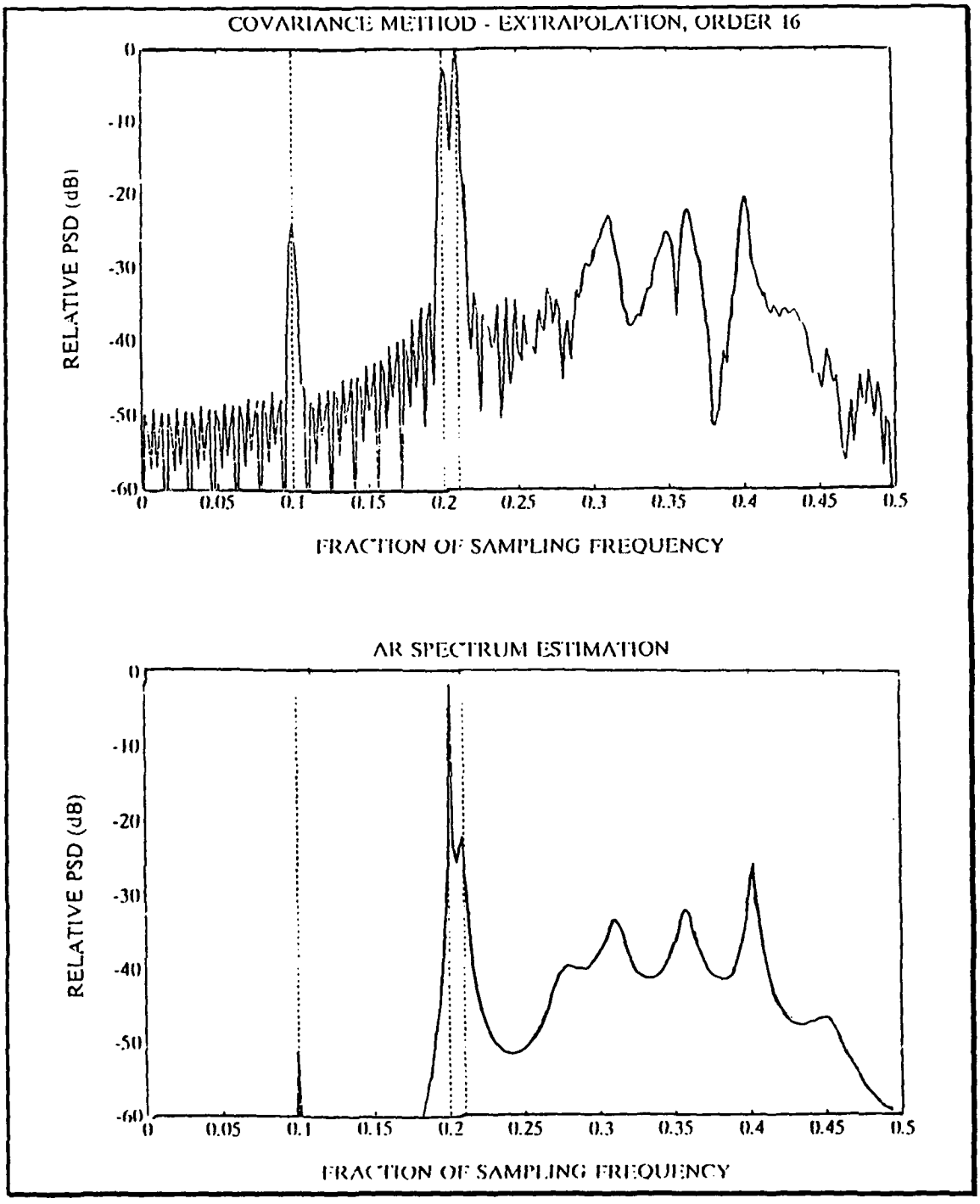

Figure 45. Periodogram of Covariance Method Extrapolated KM Data and Respective AR Spectrum Estimate

When the covariance method extrapolation was applied to a data sequence consisting of 2 sine waves embedded in white noise (SNR=15 dB), the resulting periodogram was unable to distinguish the two closely spaced sinusoids (5 Hz apart). The AR spectrum estimation also failed. Figure 46 shows the results.

C. BURG'S METHOD

In contrast to the autocorrelation and covariance methods which estimate the AR parameters directly, the Burg method estimates the reflection coefficients and then uses the Levinson recursion to obtain the AR parameter estimates.

Burg's optimization criterion is to minimize both the forward prediction error $e_P^+(n)$ and the backward prediction error $e_P^-(n)$:

$$\rho = \frac{1}{N-P}\sum_{n=P}^{N-1}\left[e_P^+[n]^2 + e_P^-[n]^2\right] \tag{A.8}$$

where

$$e_P^+[n] = x[n] + a[1]x[n-1] + a[2]x[n-2] + \ldots + a[P]x[n-P]$$
$$e_P^-[n] = x[n-P] + a[1]x[n-P+1] + a[2]x[n-P+2] + \ldots + a[P]x[n].$$

The computional steps are summarized below [Ref. 2:pp. 228-231]

1. Initialize by setting

$$e_0^+[n] = e_0^-[n] = x[n] \quad \text{for } 0 \le n \le N-1$$

and the 0th order mean square error as

$$E_0 = \frac{1}{N}\sum_{n=0}^{N-1} x[n]^2. \tag{A.9}$$

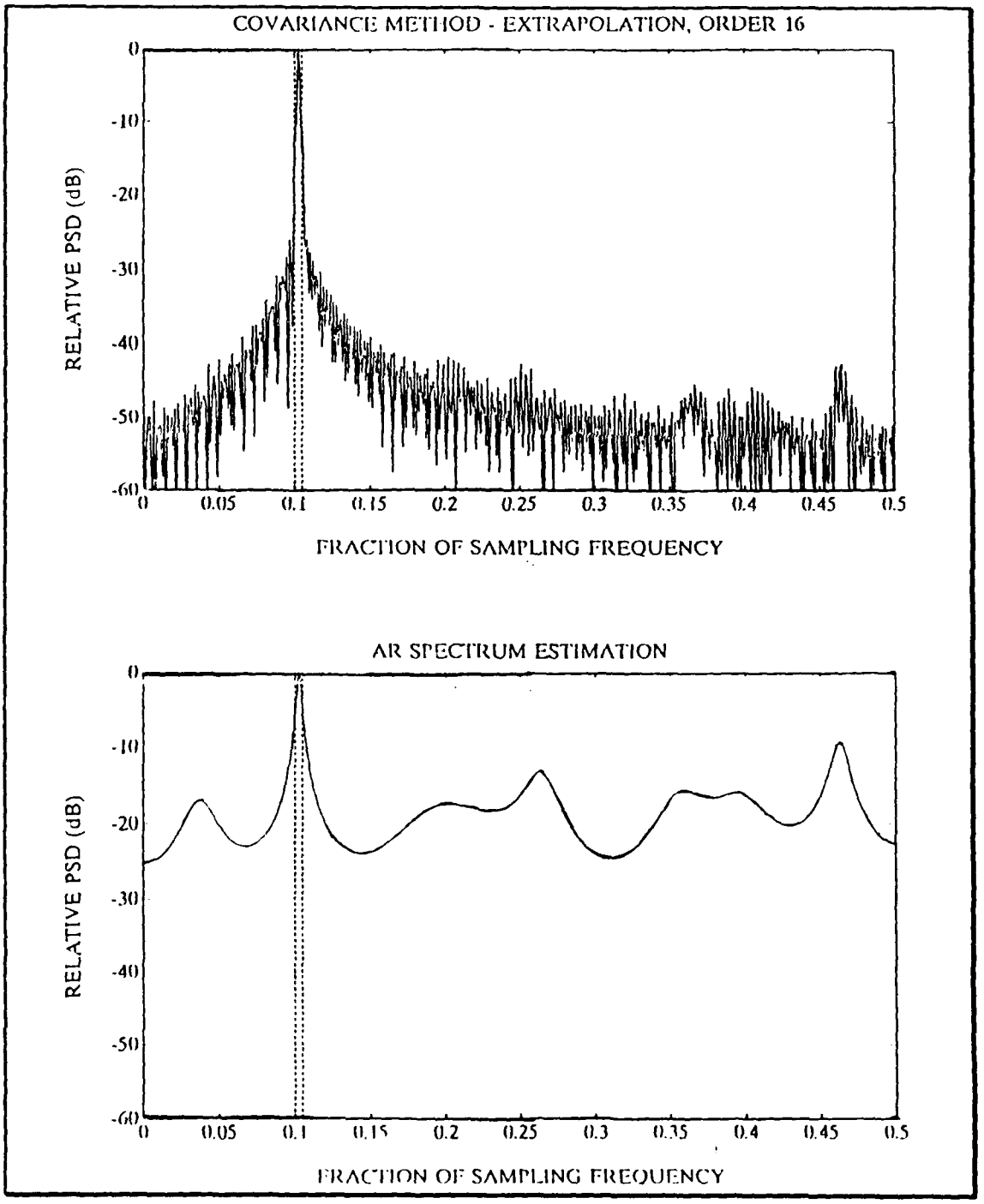

Figure 46. Periodogram of Extrapolated Sequence and Respective AR Spectrum Estimate Using Covariance Method

2. Compute the reflection coefficient:

$$\gamma_P = \frac{2\sum_{n=P}^{N-1} e^+_{P-1}[n]e^-_{P-1}[n-1]}{\sum_{n=P}^{N-1} e^+_{P-1}[n]^2 + e^-_{P-1}[n-1]^2}. \tag{A.10}$$

3. Compute $a_P[k]$ using the Levinson recursion given by:

$$\begin{bmatrix} 1 \\ a_P[1] \\ a_P[2] \\ \vdots \\ a_P[P] \end{bmatrix} = \begin{bmatrix} 1 \\ a_{P-1}[1] \\ a_{P-1}[2] \\ \vdots \\ a_{P-1}[P-1] \\ 0 \end{bmatrix} - \gamma_P \begin{bmatrix} 0 \\ a_{P-1}[P-1] \\ a_{P-1}[P-2] \\ \vdots \\ a_{P-1}[1] \\ 1 \end{bmatrix}. \tag{A.11}$$

4. Compute $e^+_P[n]$ and $e^-_P[n]$ for $P \le n \le N-1$

$$\begin{aligned} e^+_P[n] &= e^+_{P-1}[n] - \gamma_P e^-_{P-1}[n] \\ e^-_P[n] &= e^-_{P-1}[n-1] - \gamma_P e^+_{P-1}[n]. \end{aligned} \tag{A.12}$$

5. Update the mean-square error as follows

$$E_P = (1-\gamma_P)^2 E_{P-1}.$$

6. Continue the iteration until P equals the model order.

Figure 47 shows the results using Burg's method. The data is extrapolated to 320 points. The periodogram of the extrapolated KM data resolves the 3 sinusoidal components, however one of the spectral components is slightly off. In addition it seems to be prone to line splitting. The respective AR spectral estimation has similar results. The model order is 16.

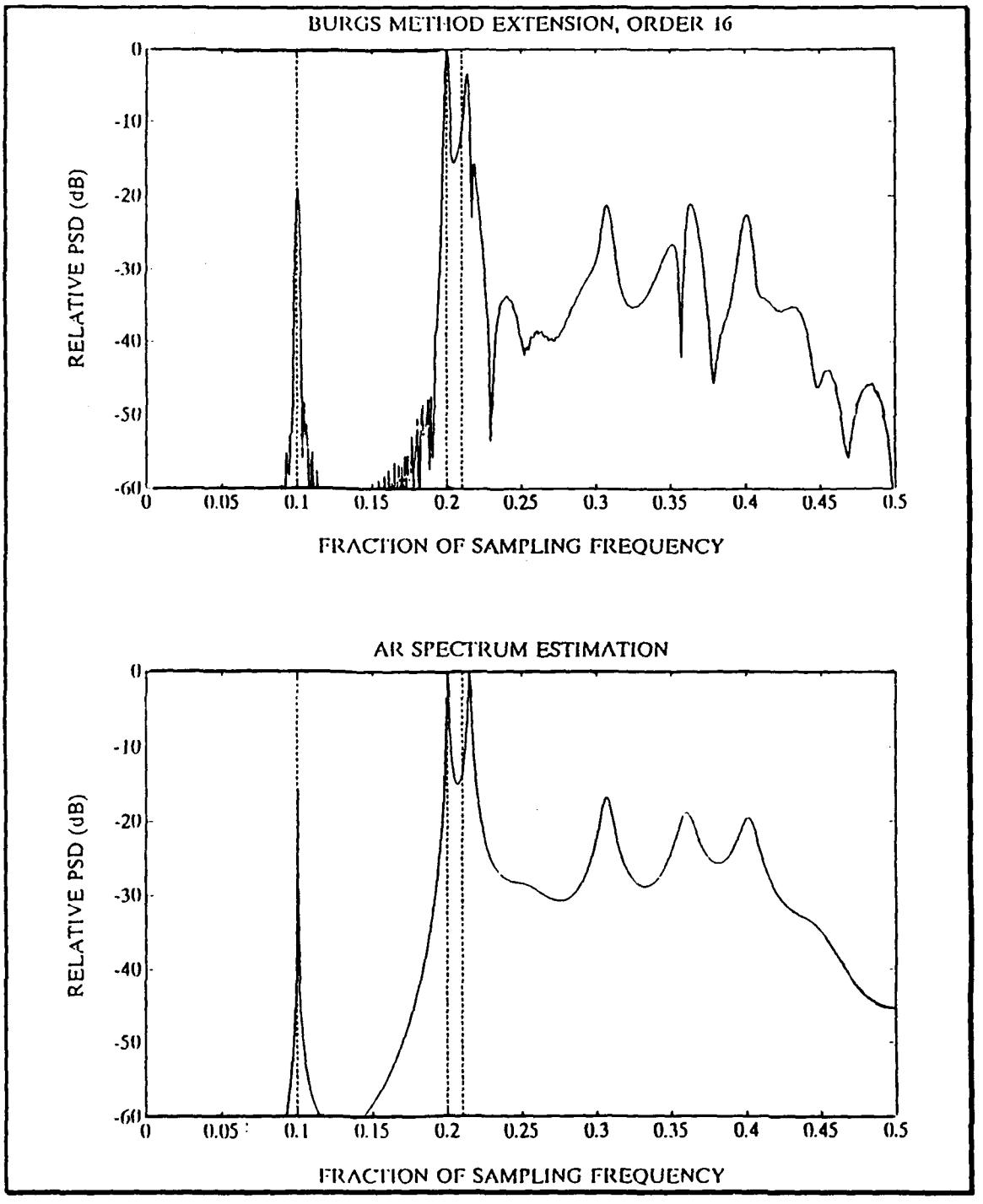

Figure 47. Periodogram of Extrapolated Sequence and Respective AR Spectrum Estimate of KM Data Using Burg's Method

Figure 48 shows the PSDs of a sequence consisting of 2 sine waves embedded in white noise (SNR=15 dB) which are 5 Hz apart (f_S =1000 Hz). Neither the periodogram of the extrapolated sequence nor the AR spectrum estimation could distinguish the two closely spaced sinusoids. The model order once again is 16 and the data is extrapolated to 320 points.

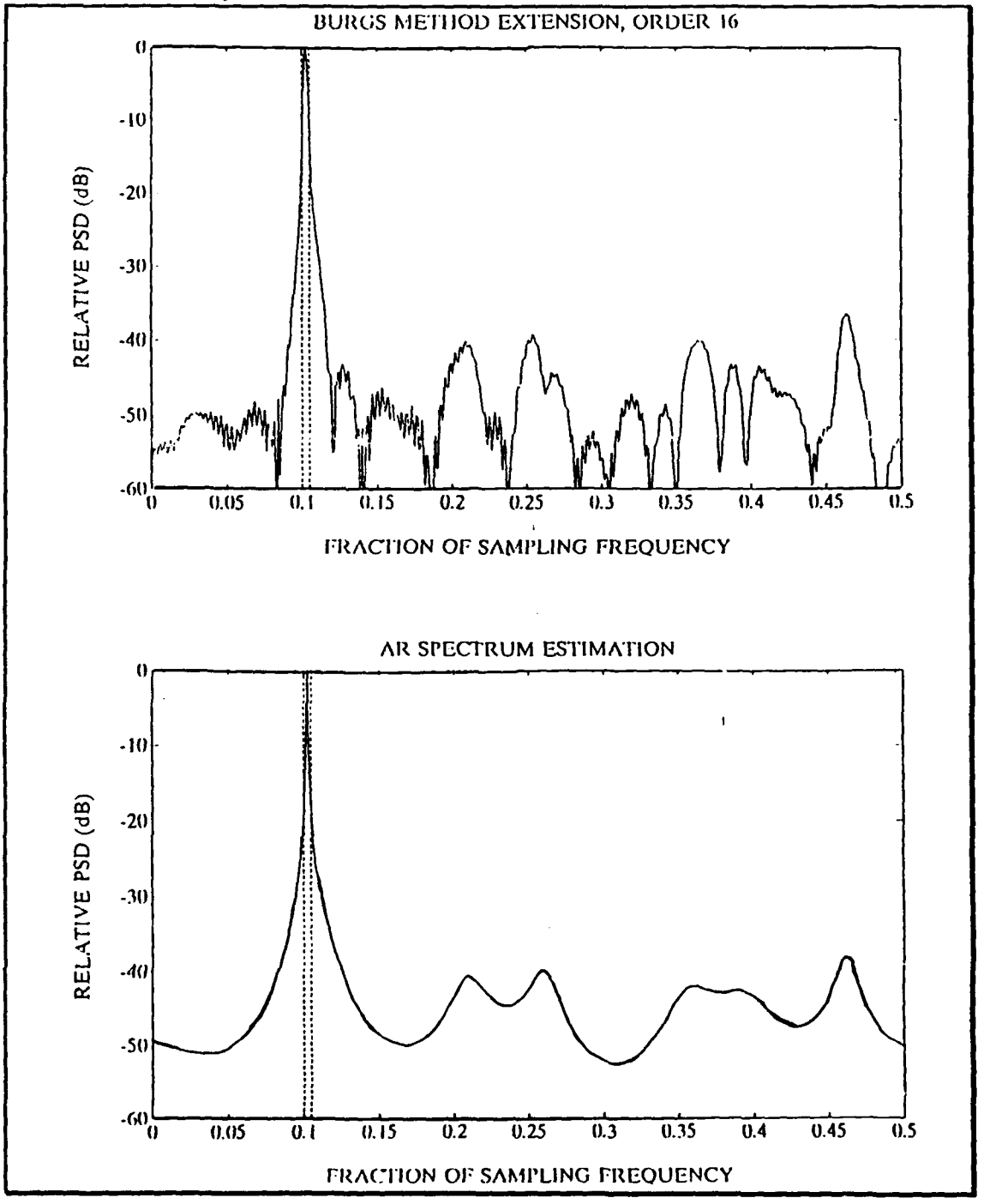

Figure 48. Periodogram of Extrapolated Sequence and Respective AR Spectrum Estimate of 2 Sine Waves using Burg's Method

APPENDIX B. SIMULATION RESULTS FROM PRONY'S METHOD

In Chapter III, the model order for the numerator and denominator in Eq. 3.2 was chosen to be 16. If too low an order is chosen, there is not enough resolution in the periodogram of the resulting extrapolated sequence. Figure 49 shows the results of a noncausal extrapolation and the resulting periodogram for the KM data when the model order is 8. The spectral component at 0.1 fraction of the sampling frequency is not resolved and the two closely spaced components at 0.2 and 0.21 are incorrect. The data extrapolation length is noted on each figure.

If the order is too high, spurious spectral peaks arise. This can be seen in Figure 50 which is the noncausal extrapolation and resulting periodogram of two sine waves embedded in white noise (SNR=15 dB) which are 10 Hz apart (f_S = 1000 Hz). The two frequency components are properly represented but an incorrect spectral component at approximately 0.25 arises.

When a causal or strictly forward extrapolation was tried, the results were not as promising as the noncausal extrapolation. Figure 51 shows the results of a causal extrapolation and the resulting periodogram of the KM data. The model order is 16. The periodogram of the causal extrapolation was unable to resolve the two closely spaced sinusoids at 0.2 and 0.21. Figure 52 shows the results of a causal extrapolation and the resulting periodogram of two sine waves embedded in white noise (SNR=15 dB) and 10 Hz apart (f_S = 1000 Hz). The periodogram from the causal extrapolation was unable to resolve the two frequencies.

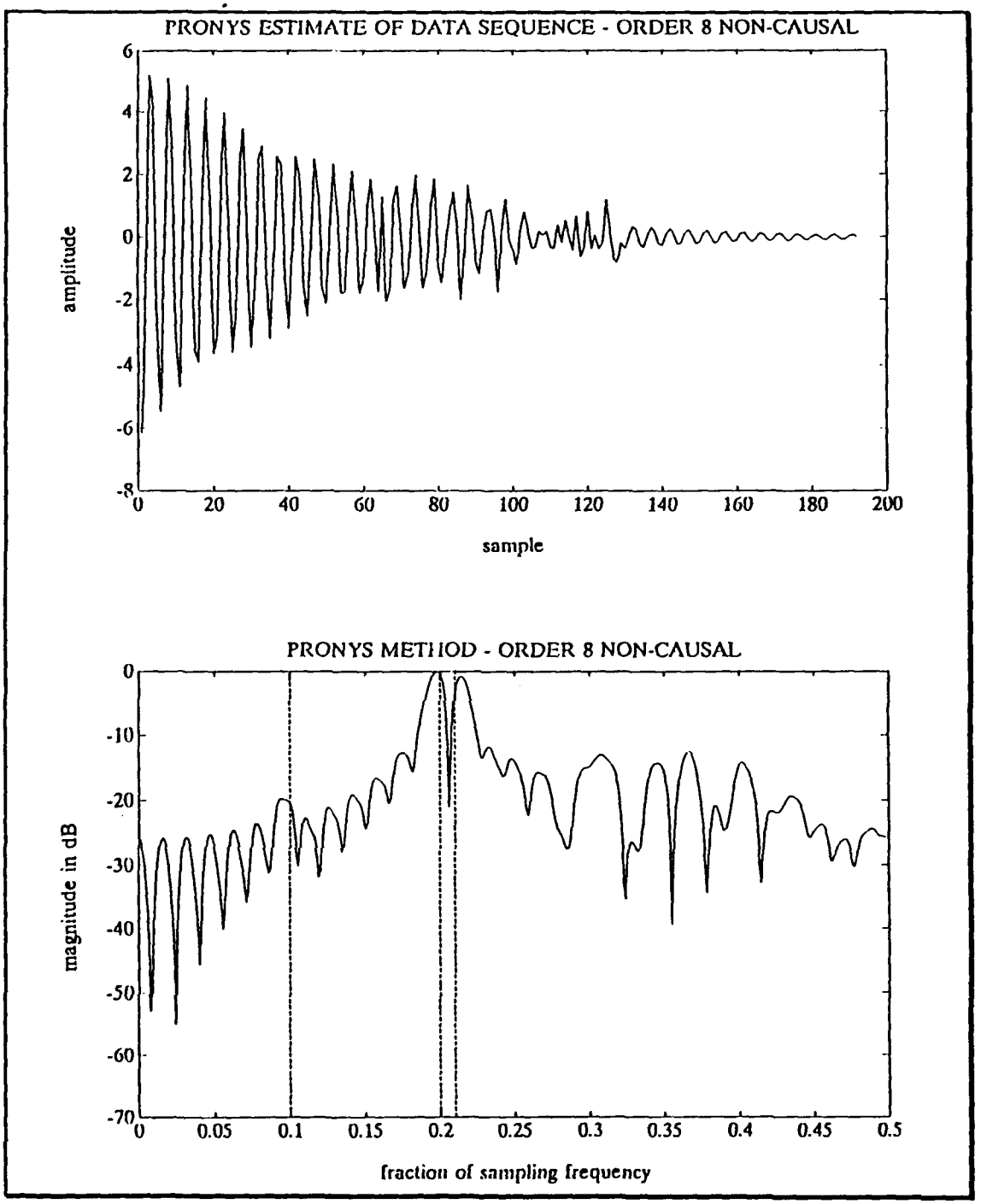

Figure 49. Noncausal Extrapolation of KM Data and Resulting Periodogram

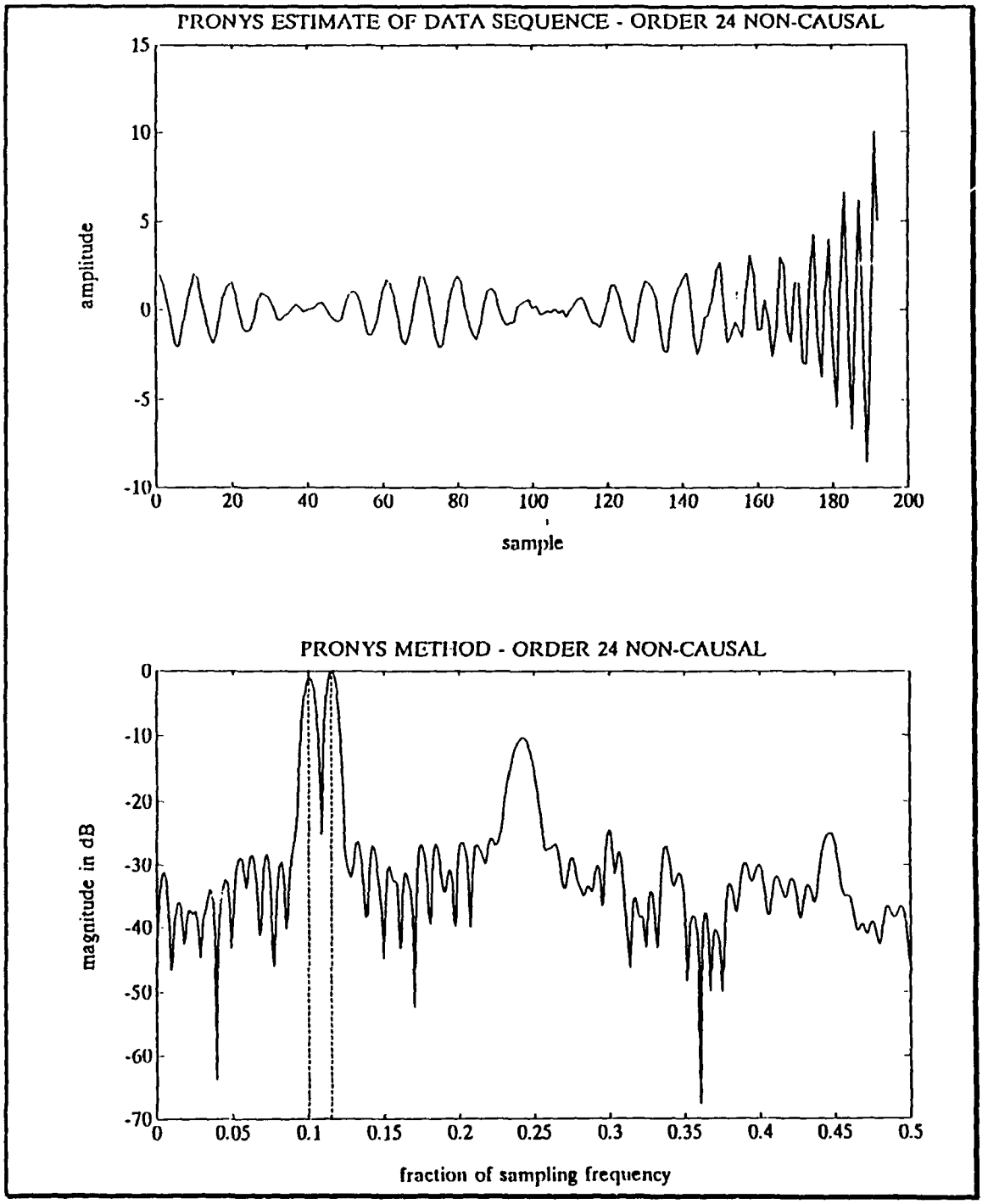

Figure 50. Noncausal Extrapolation and Resulting Periodogram of 2 Sine Waves

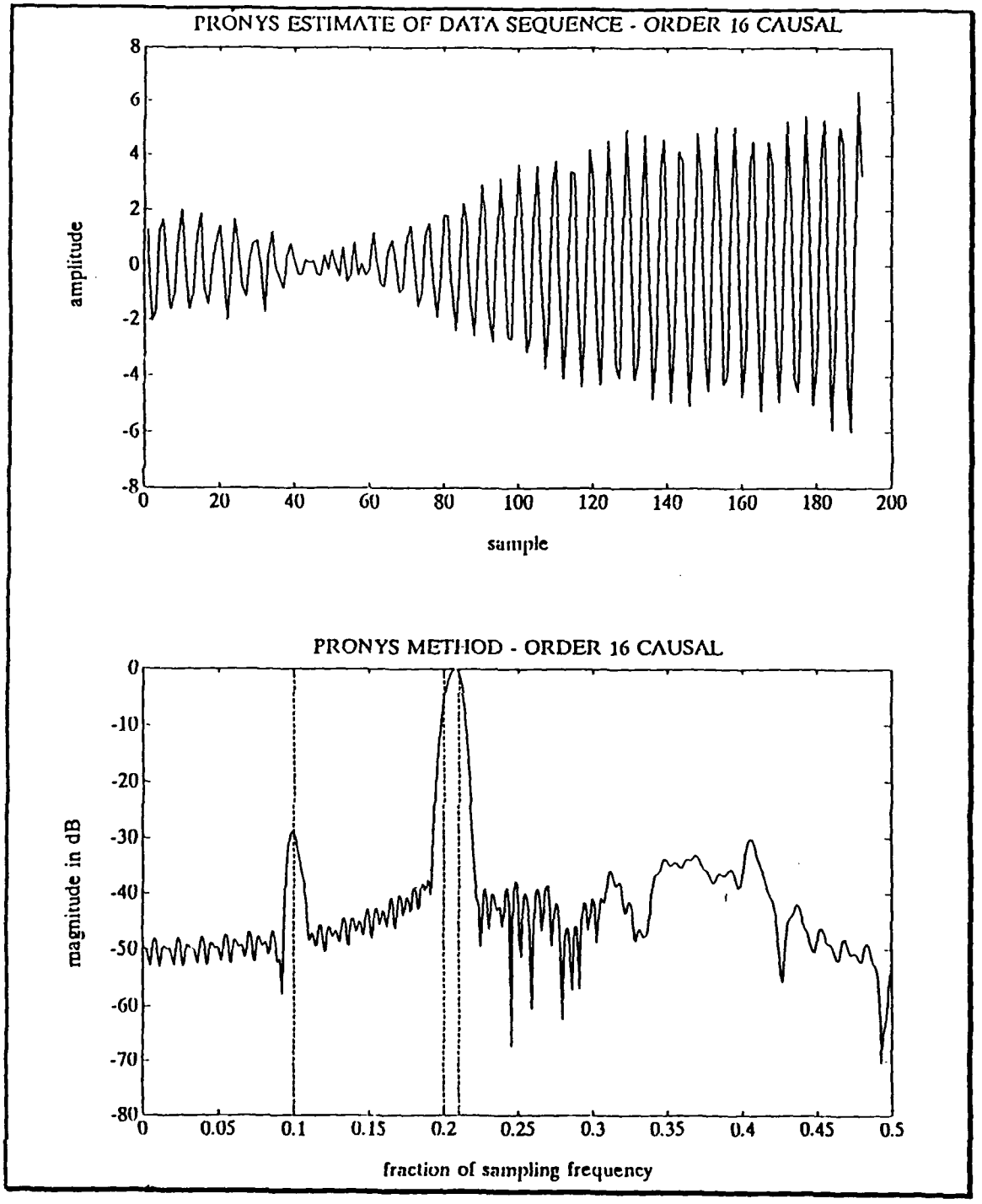

Figure 51. Causal Extrapolation and Resulting Periodogram of KM Data

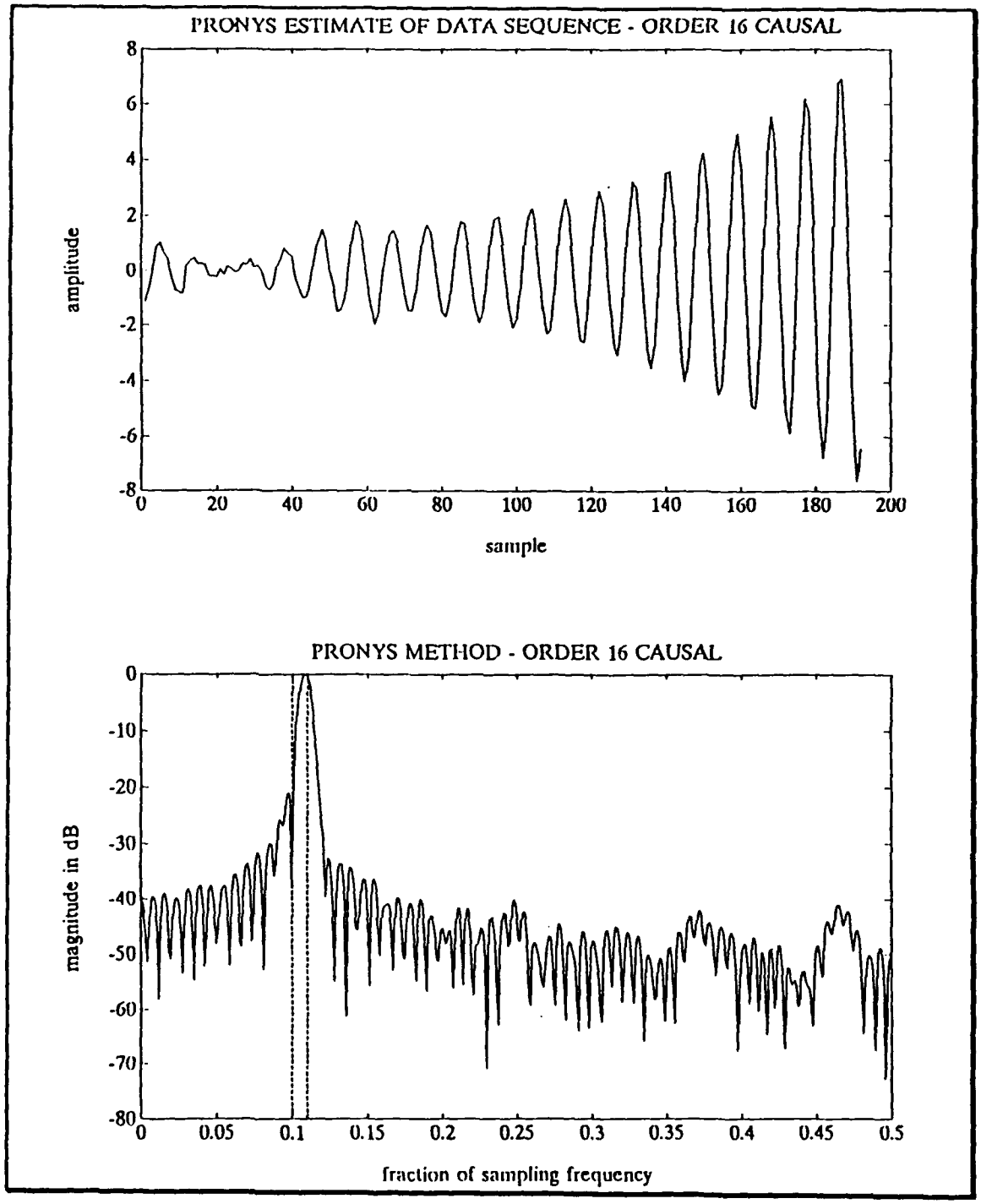

Figure 52. Causal Extrapolation and Resulting Periodogram of two Sine Waves

APPENDIX C. RESULTS FROM MISCELLANEOUS SEQUENCES

The four main types of extrapolation techniques in this thesis (modified covariance, modified forward-backward, Prony, and eigenvector method) were tried on a noise only sequence. Figures 53 through 56 show the respective results of the periodogram of the extrapolated sequences and the AR spectrum estimations. The noise sequences are extrapolated to 320 points prior to the calculation of the periodogram. Figure 57 shows the periodogram of the nonextrapolated noise sequence.

The MFBLP method and Prony's method were also tried on a test sequence consisting of a single exponentially decaying sine wave with an SNR of 15 dB. The sine wave decayed to 10% of its maximum value within the original 64 points. Figure 58 shows the resulting periodogram from the MFBLP method extrapolated sequence and the respective AR spectrum estimation. Both methods correctly identified the frequency of the sine wave. Figure 59 shows the Prony's method extrapolated sequence and the resulting periodogram. Once again the periodogram correctly identified the frequency present.

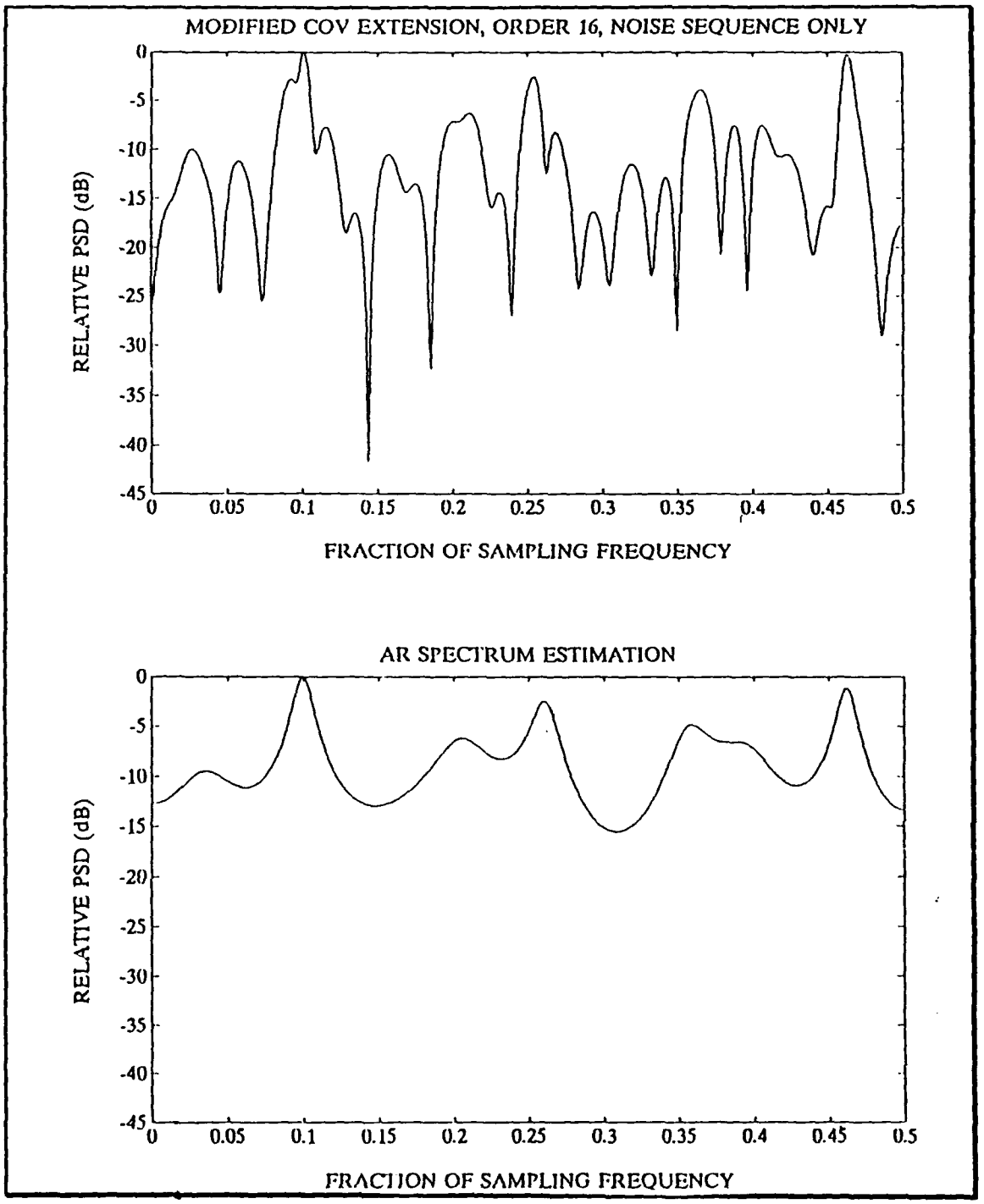

Figure 53. Periodogram of FBLP Extrapolation of Noise and Respective AR Spectrum Estimate

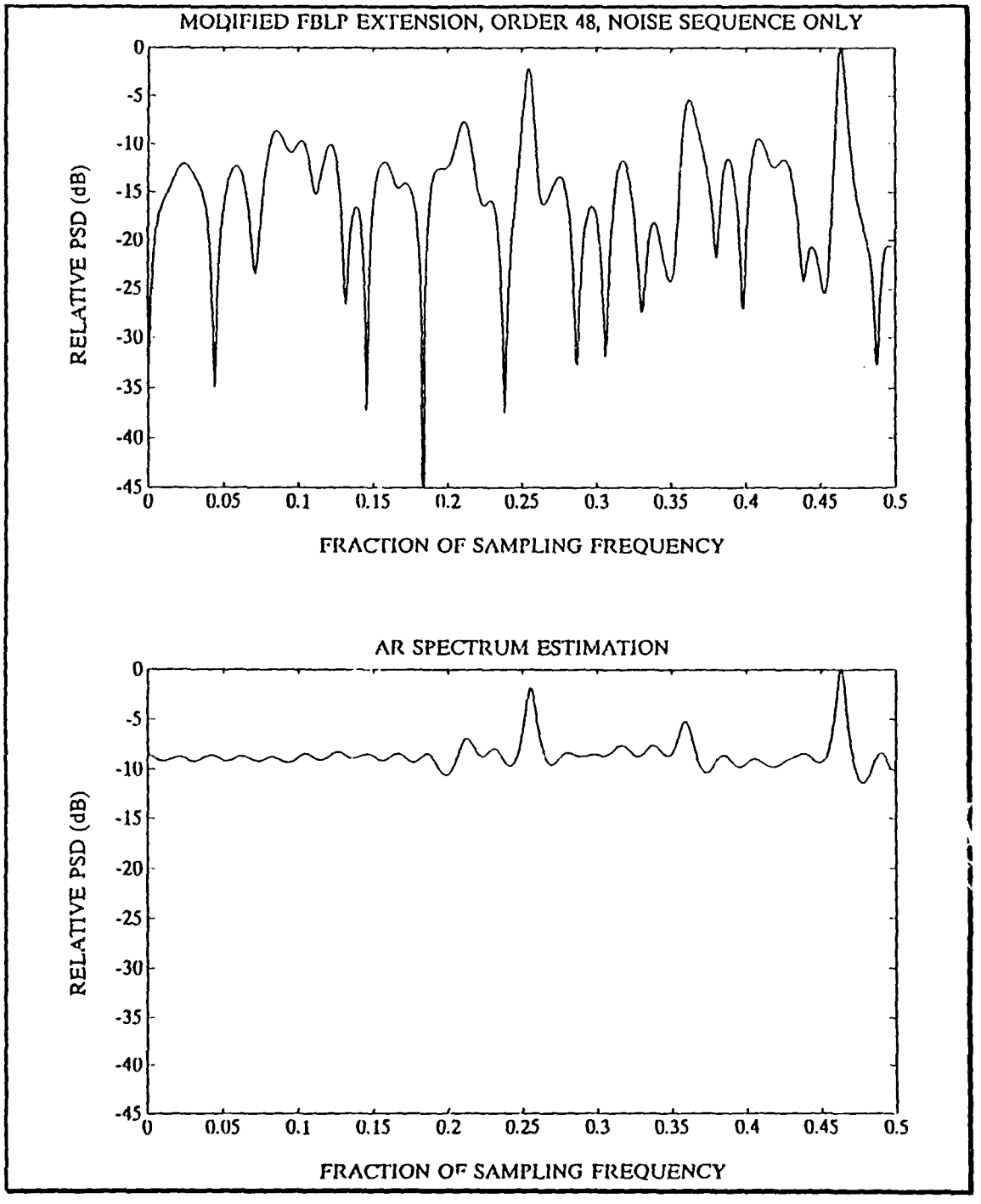

Figure 54. Periodogram of MFBLP Extrapolation of Noise and Respective AR Spectrum Estimate

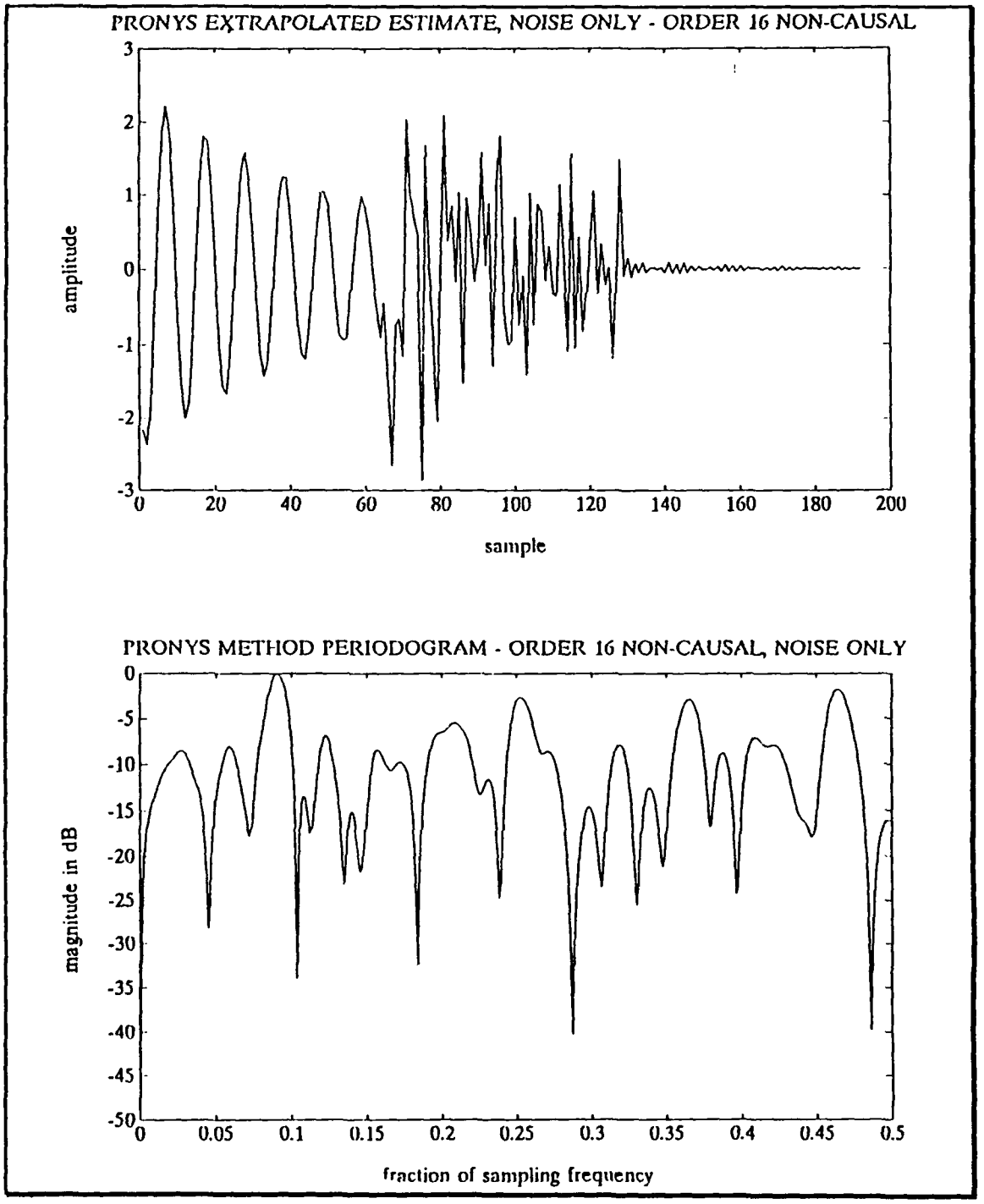

Figure 55. Prony's Extrapolation of Noise and Resulting Periodogram

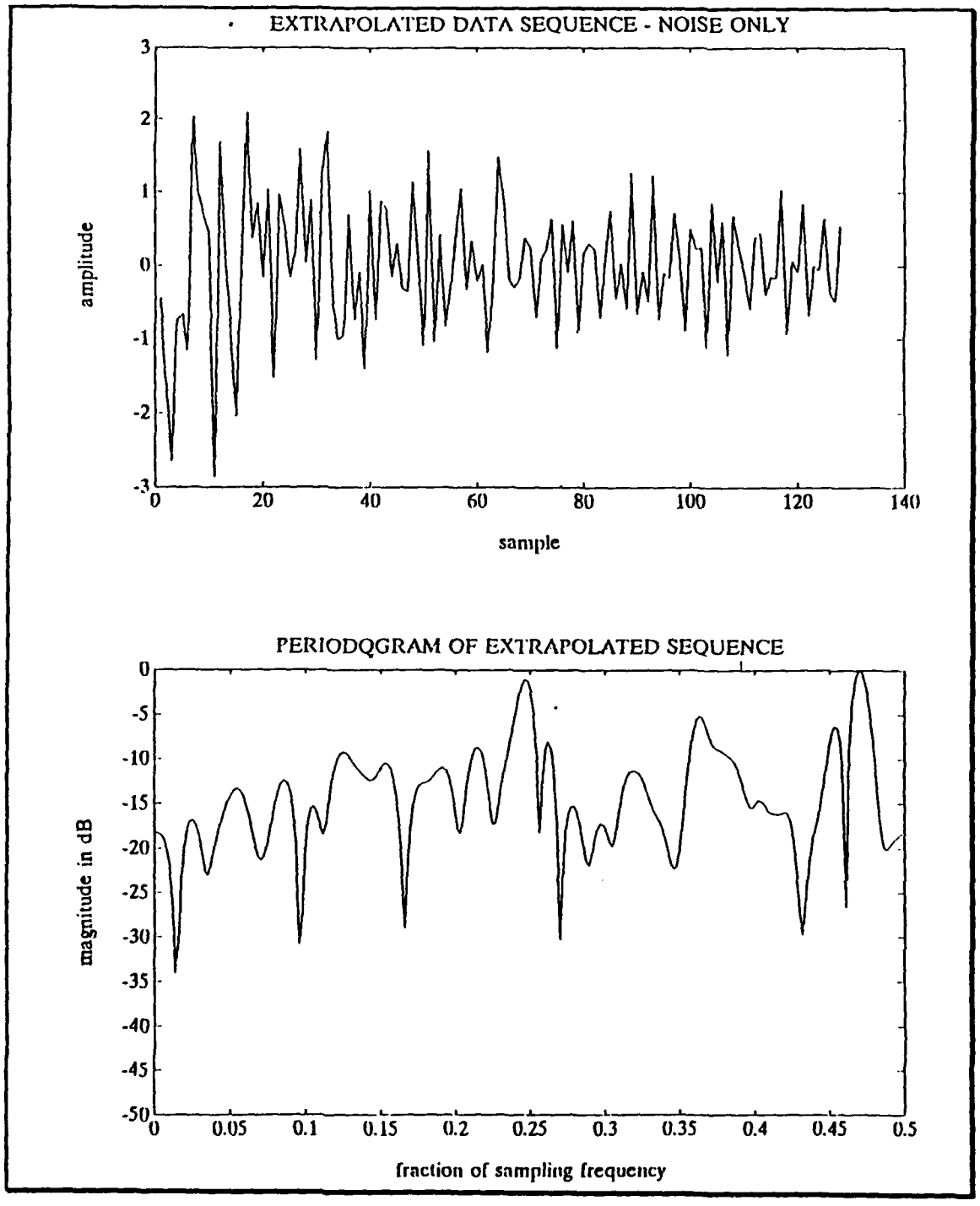

Figure 56. Eigenvector Extrapolation of Noise and Resulting Periodogram

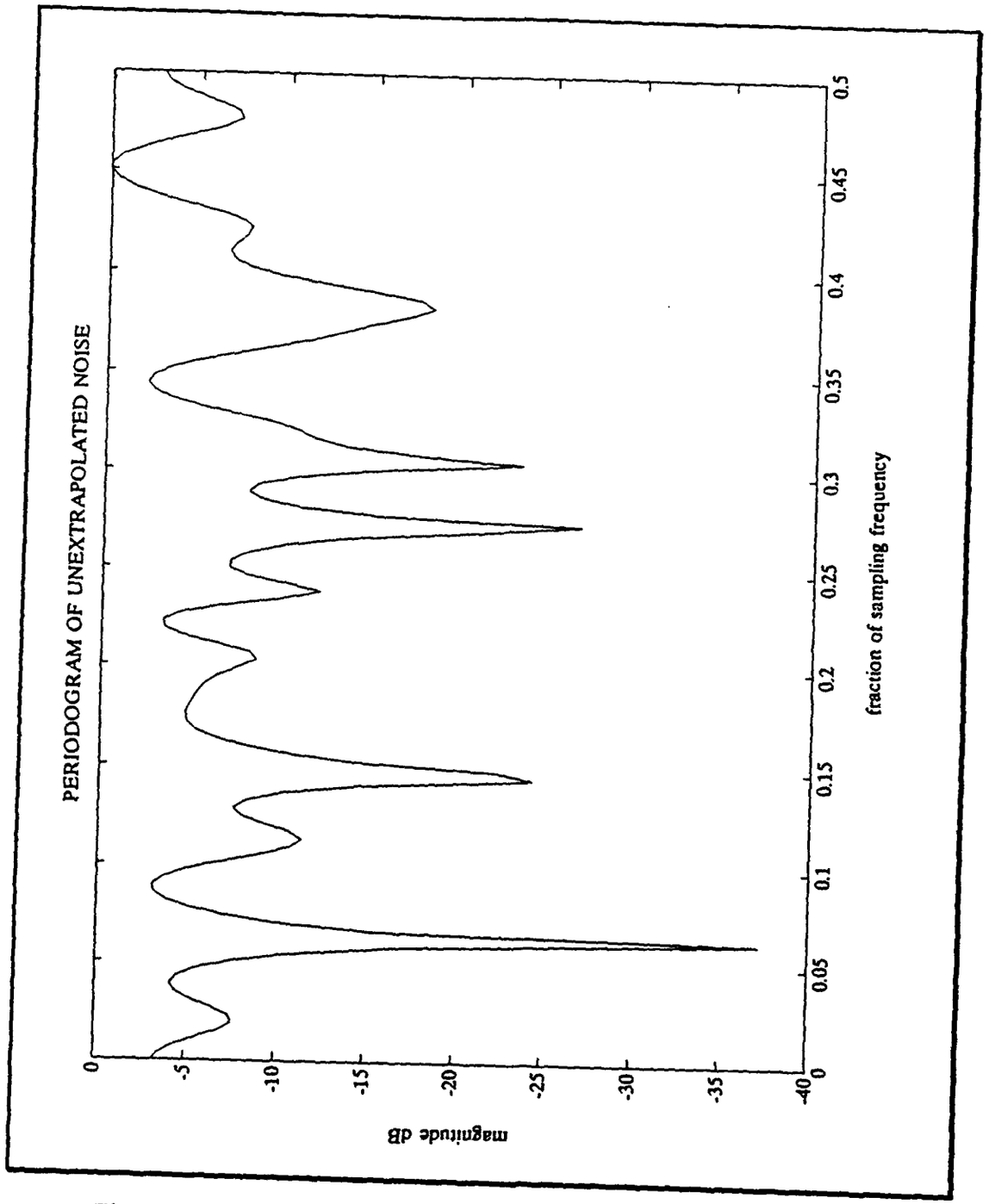

Figure 57. Periodogram of the Nonextrapolated Noise Sequence

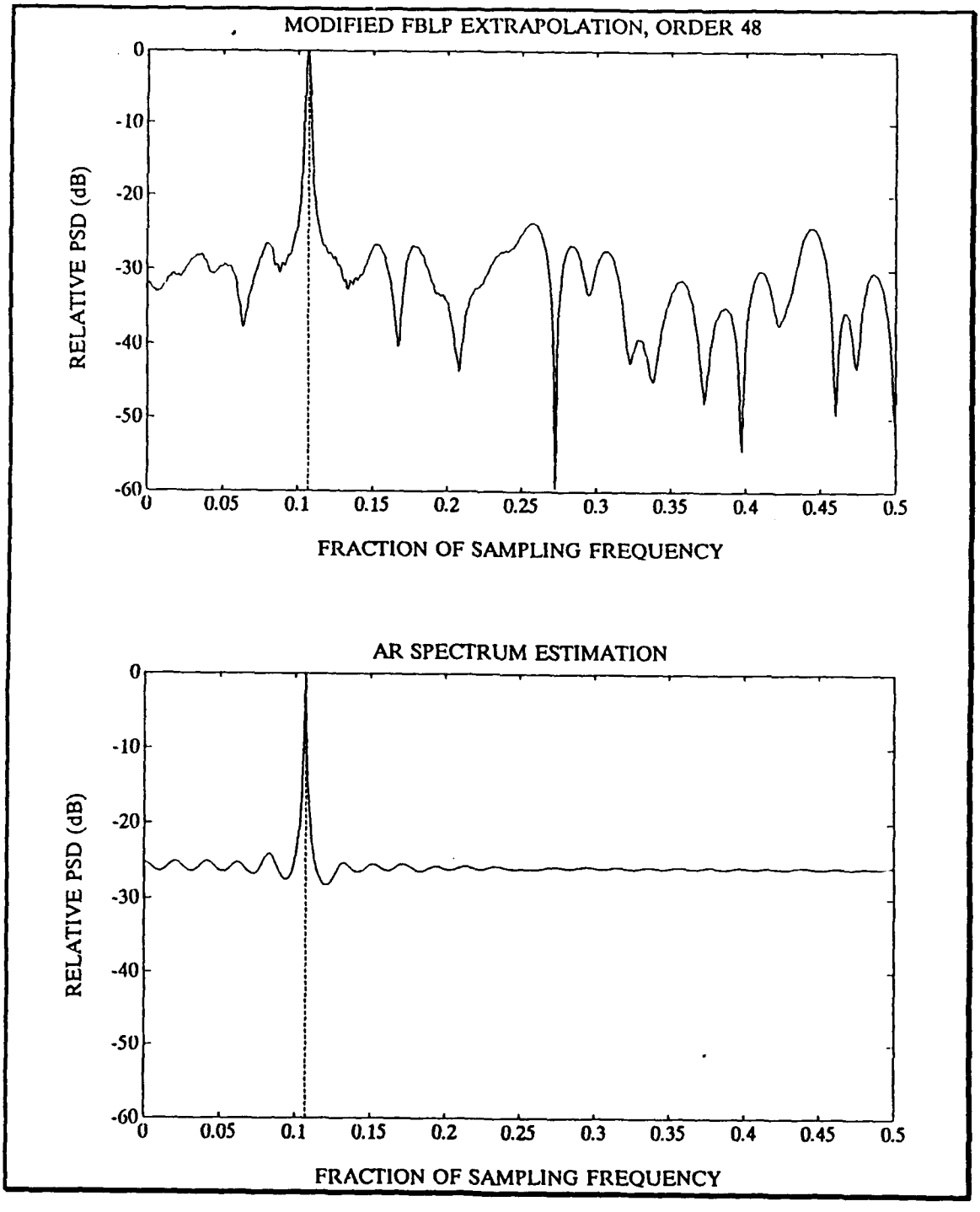

Figure 58. Periodogram of MFBLP Extrapolated Exponentially Decaying Sine Wave and Respective AR Spectrum Estimation

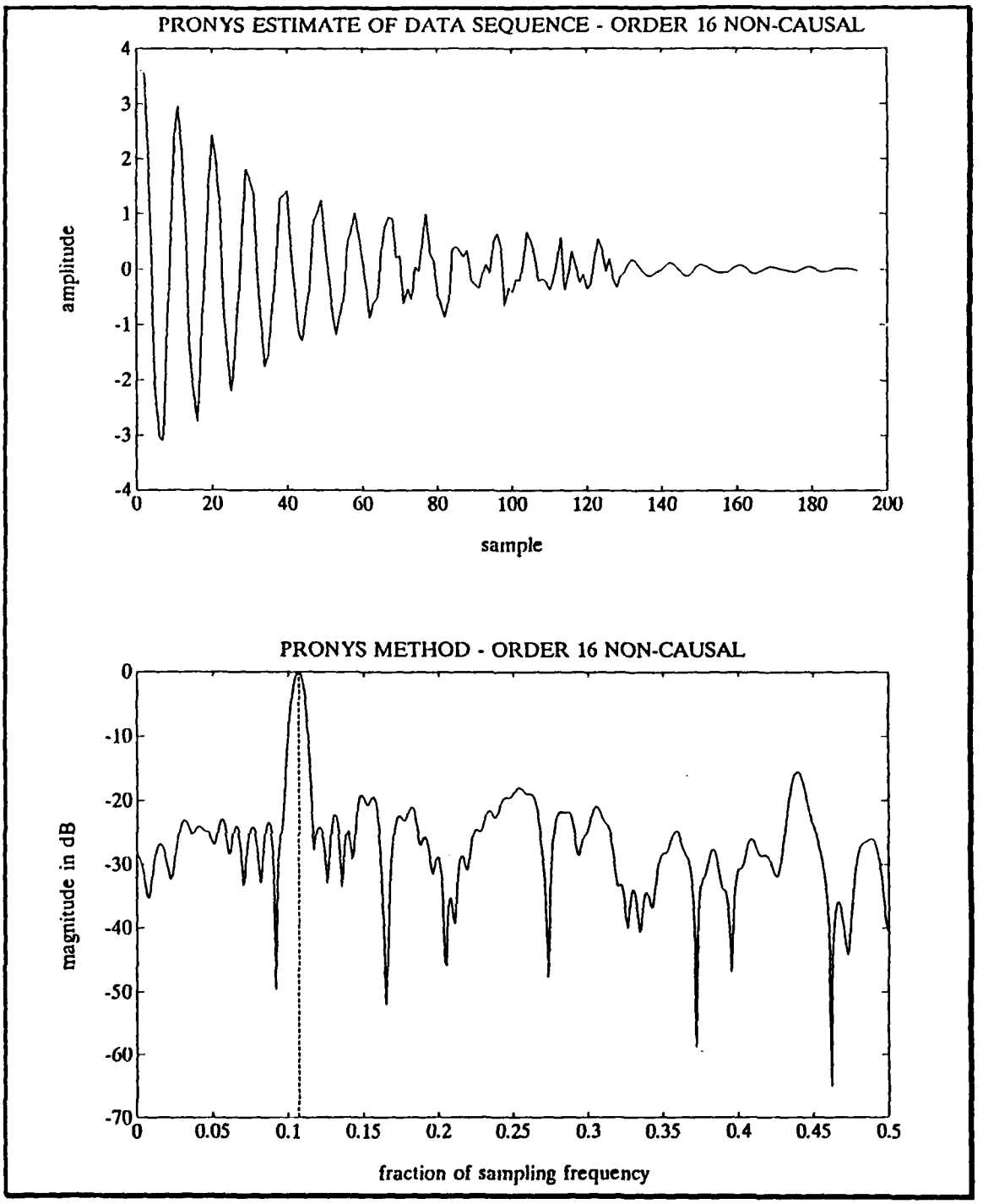

Figure 59. Prony's Extrapolation of Exponentially Decaying Sine Wave and Respective AR Spectrum Estimation

APPENDIX D. COMPUTER SIMULATION CODE

```
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%%  backfor.m
%%
%%  This program extrapolates the data using forward-
%%  backward linear prediction.  The AR spectral
%%  estimation is also calculated.
%%
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%

load kmdat.mat;                    % load data sequence
%load robd.mat;
load coeff.mat;                    % load prediction/AR
                                   % parameters
a=coeff;
%predf=robd;
predf=kmdat;

freq1=.1;
freq2=.2;
freq3=.21;

ord=16;                            % model order
ext=256;                           % amount extrapolated
pad=1024;

% Forward prediction

n=65;
for i=1:ext;
    p(i)=linpred(predf,n,a,ord);
    predf(n)=p(i);
    n=n+1;
end

% Backward prediction

flip=flipud(kmdat);                % extrapolate backwards
m=65;
for i=1:ext;
    p(i)=linpred(flip,m,a,ord);
    flip(m)=p(i);
    m=m+1;
end

predb=flipud(flip);
```

```
total=[predb(1:ext);predf];

clear predf predb flip p a kmdat

%plot(total),title('EXTRAPOLATED SEQUENCE')
%xlabel('sample'), ylabel('amplitude')
%meta bflp
%pause

k=length(total)
window=hamming(k);
finham=window .* total;
clear window total
pow=fft(finham,pad);
n=length(pow);
pyy=pow .* conj(pow) ./n;

f=0.5*(0:n/2)/(n/2);
pyy(n/2+2:n)=[];
pownorm=(1/max(pyy)) .* pyy;
powlog=10 .* log10(pownorm);
clear  pow pyy pownorm k

f1=[freq1 -100; freq1 100];
f2=[freq2 -100; freq2 100];
f3=[freq3 -100; freq3 100];
axis([0 .5 -60 0]);

a=[1;coeff];
sar=abs(fft(a,pad)) .^2;
sar= 1.0 ./ sar;
sarnorm=(1/max(sar)) .* sar;
sarlog= 10 .* log10(sarnorm);
sarlog=sarlog(1:n/2 + 1);

plot(f,powlog,'-',f1(:,1),f1(:,2),'--',f2(:,1),f2(:,2))
title('PERIODOGRAM OF EXTRAPOLATED SEQUENCE')
ylabel('RELATIVE PSD (dB)')
xlabel('FRACTION OF SAMPLING FREQUENCY')
meta bflp3
pause
plot(f,sarlog,'-',f1(:,1),f1(:,2),'--',f2(:,1),f2(:,2))
title('AR SPECTRUM ESTIMATION')
xlabel('FRACTION OF SAMPLING FREQUENCY')
ylabel('RELATIVE PSD (dB)')
meta
pause
axis;
```

```
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
function lp=linpred(x,n,a,ord)
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%

%  linpred computes the nth element of a n-1 sequence using
%  simple linear prediction, x-data sequence; n=element to be
%  predicted; a=filter coefficients; ord=order of predictor.

i=1;
for k=1:ord;
    y(i)=a(k)*x(n-k);
    i=i+1;
end

lp=-sum(y);
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%%  mfblp.m
%%
%%  This program computes the prediction/AR parameters
%%  using the modified forward/backward method.
%%
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%

%load kmdat.mat;
load robd.mat;

%x=kmdat;
x=robd;

n=length(x);                    % number of data points
m=96;
np=n-m;

for i=1:np;                     % load data matrix A
    for k=1:m;
        A(i,k)=x(i+m-k);
        A(i+np,k)=x(i+k);
    end
end

%[u,s,v]=svd(A);                % singular value decomposition
[eigvec,eigval]=eig(A'*A);



k=4;                            % number of sinusoids present
                                % in signal, assumed known
```

```
eigval=diag(eigval);
%eigval=diag(s) .^2;              % eigenvalues of A'A
w=zeros(m,1);

for i=1:n-m;
    b(i,1)=x(m+i);                % 2(n-m)-by-1 desired
    b(i+n-m,1)=x(i);              % response vector
end

for i=1:k;
    w=w + (eigvec(:,m-i+1) ./ eigval(m-i+1)*eigvec(:,m-i+1)
end


%for i=1:k;
%    w=w + (v(:,i) ./ eigval(i) * v(:,1)' * A' * b);
%end

!erase coeff.mat
diary coeff.mat
w=-w
diary off
!q coeff.mat
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%%  pron.m
%%
%%  This program finds an IIR filter with a prescribed
%%  time domain response using Prony's Method for filter
%%  design.  The impulse response of that filter is then
%%  computed and a periodogram is computed to estimate
%%  the spectrum of the data.  The data is extrapolated
%%  by increasing the impulse duration.
%%
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%

load kmdat.mat;
h=kmdat;                          % time domain impulse response
hf=flipud(h);

nb=16;                            % order of numerator
na=16;                            % order of denominator
zpad=1024;

[b,a]=prony(h,nb,na);             % IIR filter coefficients
[d,c]=prony(hf,nb,na);
n=128;
x=[1 zeros(1,n-1)];
y=filter(b,a,x);     ynew=y(65:n)';
yf=filter(d,c,x);    ynewf=flipud(yf(65:n)');
```

REFERENCES

1. Therrien, C. W., *Statistical Signal Processing—An Introduction*, Prentice-Hall, 1991.

2. Kay, S. M., *Modern Spectral Estimation*, Prentice-Hall, 1988.

3. Haykin, S., *Adaptive Filter Theory*, Prentice-Hall, 1986.

4. Lang, S. W., and McCellan, J. H., "Frequency Estimation with Maximum Entropy Spectral Estimators," *IEEE Trans. Acoust., Speech, and Signal Processing*, v. 28, March 1980.

5. Akaike, H., "Statistical Predictor Identification," *Am. Inst. Statis. Math.*, v. 22, March 1970.

6. Kay, S. M., and Marple, S. L., "Spectrum Analysis—A Modern Perspective," *Proc. IEEE*, v. 69, November 1981.

7. Tufts, D. W., and Kumaresan, R., "Estimation of Frequencies of Multiple Sinusoids: Making Linear Prediction Perform like Maximum Likelihood," *Proc. IEEE*, v. 70, October 1982.

8. Haykin, S., *Modern Filters*, Macmillan Publishing Company, 1989.

INITIAL DISTRIBUTION LIST

1.	Defense Technical Information Center Cameron Station Alexandria, VA 22304-6145	2
2.	Library, Code 52 Naval Postgraduate School Monterey, CA 93943-5002	2
3.	Chairman, Code EC Naval Postgraduate School Monterey, CA 93943-5000	1
4.	Professor Ralph Hippenstiel, Code EC/Hi Naval Postgraduate School Monterey, CA 93943-5000	3
5.	Professor Murali Tummala, Code EC/Tu Naval Postgraduate School Monterey, CA 93943-5000	1
6.	Robert T. Thornlow 1005 Ormandy Ave. Jacksonville, NC 28546	1
7.	Professor Roberto Cristi, Code EC/Cx Naval Postgraduate School Monterey, CA 93943-5000	1
8.	Professor Charles W. Therrien, Code EC/Ti Naval Postgraduate School Monterey, CA 93943-5000	1
9.	Naval Ocean Systems Center Attn: Dr. C. E. Persons, Code 732 San Diego, CA 92152	1